1. Linear Equations

1.1. Fields

We assume that the reader is familiar with the elementary algebra of real and complex numbers. For a large portion of this book the algebraic properties of numbers which we shall use are easily deduced from the following brief list of properties of addition and multiplication. We let F denote either the set of real numbers or the set of complex numbers.

1. Addition is commutative,

$$x + y = y + x$$

for all x and y in F.

2. Addition is associative,

$$x + (y + z) = (x + y) + z$$

for all x, y, and z in F.

3. There is a unique element 0 (zero) in F such that $x + 0 = x$, for every x in F.

4. To each x in F there corresponds a unique element $(-x)$ in F such that $x + (-x) = 0$.

5. Multiplication is commutative,

$$xy = yx$$

for all x and y in F.

6. Multiplication is associative,

$$x(yz) = (xy)z$$

for all x, y, and z in F.

7. There is a unique non-zero element 1 (one) in F such that $x1 = x$, for every x in F.

8. To each non-zero x in F there corresponds a unique element x^{-1} (or $1/x$) in F such that $xx^{-1} = 1$.

9. Multiplication distributes over addition; that is, $x(y + z) = xy + xz$, for all x, y, and z in F.

Suppose one has a set F of objects x, y, z, . . . and two operations on the elements of F as follows. The first operation, called addition, associates with each pair of elements x, y in F an element $(x + y)$ in F; the second operation, called multiplication, associates with each pair x, y an element xy in F; and these two operations satisfy conditions (1)–(9) above. The set F, together with these two operations, is then called a **field.** Roughly speaking, a field is a set together with some operations on the objects in that set which behave like ordinary addition, subtraction, multiplication, and division of numbers in the sense that they obey the nine rules of algebra listed above. With the usual operations of addition and multiplication, the set C of complex numbers is a field, as is the set R of real numbers.

For most of this book the 'numbers' we use may as well be the elements from any field F. To allow for this generality, we shall use the word 'scalar' rather than 'number.' Not much will be lost to the reader if he always assumes that the field of scalars is a subfield of the field of complex numbers. A **subfield** of the field C is a set F of complex numbers which is itself a field under the usual operations of addition and multiplication of complex numbers. This means that 0 and 1 are in the set F, and that if x and y are elements of F, so are $(x + y)$, $-x$, xy, and x^{-1} (if $x \neq 0$). An example of such a subfield is the field R of real numbers; for, if we identify the real numbers with the complex numbers $(a + ib)$ for which $b = 0$, the 0 and 1 of the complex field are real numbers, and if x and y are real, so are $(x + y)$, $-x$, xy, and x^{-1} (if $x \neq 0$). We shall give other examples below. The point of our discussing subfields is essentially this: If we are working with scalars from a certain subfield of C, then the performance of the operations of addition, subtraction, multiplication, or division on these scalars does not take us out of the given subfield.

EXAMPLE 1. The set of **positive integers:** 1, 2, 3, . . . , is not a subfield of C, for a variety of reasons. For example, 0 is not a positive integer; for no positive integer n is $-n$ a positive integer; for no positive integer n except 1 is $1/n$ a positive integer.

EXAMPLE 2. The set of **integers:** . . . , -2, -1, 0, 1, 2, . . . , is not a subfield of C, because for an integer n, $1/n$ is not an integer unless n is 1 or

-1. With the usual operations of addition and multiplication, the set of integers satisfies all of the conditions (1)–(9) except condition (8).

EXAMPLE 3. The set of **rational numbers**, that is, numbers of the form p/q, where p and q are integers and $q \neq 0$, is a subfield of the field of complex numbers. The division which is not possible within the set of integers is possible within the set of rational numbers. The interested reader should verify that any subfield of C must contain every rational number.

EXAMPLE 4. The set of all complex numbers of the form $x + y\sqrt{2}$, where x and y are rational, is a subfield of C. We leave it to the reader to verify this.

In the examples and exercises of this book, the reader should assume that the field involved is a subfield of the complex numbers, unless it is expressly stated that the field is more general. We do not want to dwell on this point; however, we should indicate why we adopt such a convention. If F is a field, it may be possible to add the unit 1 to itself a finite number of times and obtain 0 (see Exercise 5 following Section 1.2):

$$1 + 1 + \cdots + 1 = 0.$$

That does not happen in the complex number field (or in any subfield thereof). If it does happen in F, then the least n such that the sum of n 1's is 0 is called the **characteristic** of the field F. If it does not happen in F, then (for some strange reason) F is called a field of **characteristic zero.** Often, when we assume F is a subfield of C, what we want to guarantee is that F is a field of characteristic zero; but, in a first exposure to linear algebra, it is usually better not to worry too much about characteristics of fields.

1.2. *Systems of Linear Equations*

Suppose F is a field. We consider the problem of finding n scalars (elements of F) $x_1, \ldots, x_n$ which satisfy the conditions

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= y_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= y_2 \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots & \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= y_m \end{aligned} \tag{1-1}$$

where $y_1, \ldots, y_m$ and A_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, are given elements of F. We call (1-1) a **system of m linear equations in n unknowns.** Any n-tuple $(x_1, \ldots, x_n)$ of elements of F which satisfies each of the

equations in (1-1) is called a **solution** of the system. If $y_1 = y_2 = \cdots = y_m = 0$, we say that the system is **homogeneous,** or that each of the equations is homogeneous.

Perhaps the most fundamental technique for finding the solutions of a system of linear equations is the technique of elimination. We can illustrate this technique on the homogeneous system

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 0 \\ x_1 + 3x_2 + 4x_3 &= 0. \end{aligned}$$

If we add (-2) times the second equation to the first equation, we obtain

$$-7x_2 - 7x_3 = 0$$

or, $x_2 = -x_3$. If we add 3 times the first equation to the second equation, we obtain

$$7x_1 + 7x_3 = 0$$

or, $x_1 = -x_3$. So we conclude that if (x_1, x_2, x_3) is a solution then $x_1 = x_2 = -x_3$. Conversely, one can readily verify that any such triple is a solution. Thus the set of solutions consists of all triples $(-a, -a, a)$.

We found the solutions to this system of equations by 'eliminating unknowns,' that is, by multiplying equations by scalars and then adding to produce equations in which some of the x_j were not present. We wish to formalize this process slightly so that we may understand why it works, and so that we may carry out the computations necessary to solve a system in an organized manner.

For the general system (1-1), suppose we select m scalars $c_1, \ldots, c_m$, multiply the jth equation by c_j and then add. We obtain the equation

$$(c_1A_{11} + \cdots + c_mA_{m1})x_1 + \cdots + (c_1A_{1n} + \cdots + c_mA_{mn})x_n = c_1y_1 + \cdots + c_my_m.$$

Such an equation we shall call a **linear combination** of the equations in (1-1). Evidently, any solution of the entire system of equations (1-1) will also be a solution of this new equation. This is the fundamental idea of the elimination process. If we have another system of linear equations

$$\begin{aligned} B_{11}x_1 + \cdots + B_{1n}x_n &= z_1 \\ \vdots \qquad\qquad & \quad \vdots \quad \vdots \\ B_{k1}x_1 + \cdots + B_{kn}x_n &= z_k \end{aligned} \tag{1-2}$$

in which each of the k equations is a linear combination of the equations in (1-1), then every solution of (1-1) is a solution of this new system. Of course it may happen that some solutions of (1-2) are not solutions of (1-1). This clearly does not happen if each equation in the original system is a linear combination of the equations in the new system. Let us say that two systems of linear equations are **equivalent** if each equation in each system is a linear combination of the equations in the other system. We can then formally state our observations as follows.

Theorem 1. *Equivalent systems of linear equations have exactly the same solutions.*

If the elimination process is to be effective in finding the solutions of a system like (1-1), then one must see how, by forming linear combinations of the given equations, to produce an equivalent system of equations which is easier to solve. In the next section we shall discuss one method of doing this.

Exercises

1. Verify that the set of complex numbers described in Example 4 is a subfield of C.

2. Let F be the field of complex numbers. Are the following two systems of linear equations equivalent? If so, express each equation in each system as a linear combination of the equations in the other system.

$$\begin{aligned} x_1 - x_2 &= 0 \\ 2x_1 + x_2 &= 0 \end{aligned} \qquad \begin{aligned} 3x_1 + x_2 &= 0 \\ x_1 + x_2 &= 0 \end{aligned}$$

3. Test the following systems of equations as in Exercise 2.

$$\begin{aligned} -x_1 + x_2 + 4x_3 &= 0 \\ x_1 + 3x_2 + 8x_3 &= 0 \\ \tfrac{1}{2}x_1 + x_2 + \tfrac{5}{2}x_3 &= 0 \end{aligned} \qquad \begin{aligned} x_1 - x_3 &= 0 \\ x_2 + 3x_3 &= 0 \end{aligned}$$

4. Test the following systems as in Exercise 2.

$$\begin{aligned} 2x_1 + (-1+i)x_2 + x_4 &= 0 \\ 3x_2 - 2ix_3 + 5x_4 &= 0 \end{aligned} \qquad \begin{aligned} \left(1 + \frac{i}{2}\right)x_1 + 8x_2 - ix_3 - x_4 &= 0 \\ \tfrac{2}{3}x_1 - \tfrac{1}{2}x_2 + x_3 + 7x_4 &= 0 \end{aligned}$$

5. Let F be a set which contains exactly two elements, 0 and 1. Define an addition and multiplication by the tables:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

Verify that the set F, together with these two operations, is a field.

6. Prove that if two homogeneous systems of linear equations in two unknowns have the same solutions, then they are equivalent.

7. Prove that each subfield of the field of complex numbers contains every rational number.

8. Prove that each field of characteristic zero contains a copy of the rational number field.

1.3. Matrices and Elementary Row Operations

One cannot fail to notice that in forming linear combinations of linear equations there is no need to continue writing the 'unknowns' $x_1, \ldots, x_n$, since one actually computes only with the coefficients A_{ij} and the scalars y_i. We shall now abbreviate the system (1-1) by

$$AX = Y$$

where

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

We call A the **matrix of coefficients** of the system. Strictly speaking, the rectangular array displayed above is not a matrix, but is a representation of a matrix. An $m \times n$ **matrix over the field** F is a function A from the set of pairs of integers (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, into the field F. The **entries** of the matrix A are the scalars $A(i, j) = A_{ij}$, and quite often it is most convenient to describe the matrix by displaying its entries in a rectangular array having m rows and n columns, as above. Thus X (above) is, or defines, an $n \times 1$ matrix and Y is an $m \times 1$ matrix. For the time being, $AX = Y$ is nothing more than a shorthand notation for our system of linear equations. Later, when we have defined a multiplication for matrices, it will mean that Y is the product of A and X.

We wish now to consider operations on the rows of the matrix A which correspond to forming linear combinations of the equations in the system $AX = Y$. We restrict our attention to three **elementary row operations** on an $m \times n$ matrix A over the field F:

1. multiplication of one row of A by a non-zero scalar c;
2. replacement of the rth row of A by row r plus c times row s, c any scalar and $r \neq s$;
3. interchange of two rows of A.

An elementary row operation is thus a special type of function (rule) e which associated with each $m \times n$ matrix A an $m \times n$ matrix $e(A)$. One can precisely describe e in the three cases as follows:

1. $e(A)_{ij} = A_{ij}$ if $i \neq r$, $e(A)_{rj} = cA_{rj}$.
2. $e(A)_{ij} = A_{ij}$ if $i \neq r$, $e(A)_{rj} = A_{rj} + cA_{sj}$.
3. $e(A)_{ij} = A_{ij}$ if i is different from both r and s, $e(A)_{rj} = A_{sj}$, $e(A)_{sj} = A_{rj}$.

In defining $e(A)$, it is not really important how many columns A has, but the number of rows of A is crucial. For example, one must worry a little to decide what is meant by interchanging rows 5 and 6 of a 5×5 matrix. To avoid any such complications, we shall agree that an elementary row operation e is defined on the class of all $m \times n$ matrices over F, for some fixed m but any n. In other words, a particular e is defined on the class of all m-rowed matrices over F.

One reason that we restrict ourselves to these three simple types of row operations is that, having performed such an operation e on a matrix A, we can recapture A by performing a similar operation on $e(A)$.

Theorem 2. *To each elementary row operation* e *there corresponds an elementary row operation* e_1, *of the same type as* e, *such that* $\mathrm{e}_1(\mathrm{e}(\mathrm{A})) = \mathrm{e}(\mathrm{e}_1(\mathrm{A})) = \mathrm{A}$ *for each* A. *In other words, the inverse operation (function) of an elementary row operation exists and is an elementary row operation of the same type.*

Proof. (1) Suppose e is the operation which multiplies the rth row of a matrix by the non-zero scalar c. Let e_1 be the operation which multiplies row r by c^{-1}. (2) Suppose e is the operation which replaces row r by row r plus c times row s, $r \neq s$. Let e_1 be the operation which replaces row r by row r plus $(-c)$ times row s. (3) If e interchanges rows r and s, let $e_1 = e$. In each of these three cases we clearly have $e_1(e(A)) = e(e_1(A)) = A$ for each A. ∎

Definition. *If* A *and* B *are* m × n *matrices over the field* F, *we say that* B **is row-equivalent to** A *if* B *can be obtained from* A *by a finite sequence of elementary row operations.*

Using Theorem 2, the reader should find it easy to verify the following. Each matrix is row-equivalent to itself; if B is row-equivalent to A, then A is row-equivalent to B; if B is row-equivalent to A and C is row-equivalent to B, then C is row-equivalent to A. In other words, row-equivalence is an equivalence relation (see Appendix).

Theorem 3. *If* A *and* B *are row-equivalent* m × n *matrices, the homogeneous systems of linear equations* AX = 0 *and* BX = 0 *have exactly the same solutions.*

Proof. Suppose we pass from A to B by a finite sequence of elementary row operations:

$$A = A_0 \to A_1 \to \cdots \to A_k = B.$$

It is enough to prove that the systems $A_jX = 0$ and $A_{j+1}X = 0$ have the same solutions, i.e., that one elementary row operation does not disturb the set of solutions.

So suppose that B is obtained from A by a single elementary row operation. No matter which of the three types the operation is, (1), (2), or (3), each equation in the system $BX = 0$ will be a linear combination of the equations in the system $AX = 0$. Since the inverse of an elementary row operation is an elementary row operation, each equation in $AX = 0$ will also be a linear combination of the equations in $BX = 0$. Hence these two systems are equivalent, and by Theorem 1 they have the same solutions. ∎

EXAMPLE 5. Suppose F is the field of rational numbers, and

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix}.$$

We shall perform a finite sequence of elementary row operations on A, indicating by numbers in parentheses the type of operation performed.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & -2 & -1 & 7 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & \frac{15}{2} & -\frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix}$$

The row-equivalence of A with the final matrix in the above sequence tells us in particular that the solutions of

$$\begin{aligned} 2x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 4x_2 - x_4 &= 0 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= 0 \end{aligned}$$

and

$$\begin{aligned} x_3 - \tfrac{11}{3}x_4 &= 0 \\ x_1 + \tfrac{17}{3}x_4 &= 0 \\ x_2 - \tfrac{5}{3}x_4 &= 0 \end{aligned}$$

are exactly the same. In the second system it is apparent that if we assign

any rational value c to x_4 we obtain a solution $(-\frac{17}{3}c, \frac{5}{3}, \frac{11}{3}c, c)$, and also that every solution is of this form.

EXAMPLE 6. Suppose F is the field of complex numbers and

$$A = \begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix}.$$

In performing row operations it is often convenient to combine several operations of type (2). With this in mind

$$\begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 2+i \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & 1 \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Thus the system of equations

$$\begin{aligned} -x_1 + ix_2 &= 0 \\ -ix_1 + 3x_2 &= 0 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

has only the trivial solution $x_1 = x_2 = 0$.

In Examples 5 and 6 we were obviously not performing row operations at random. Our choice of row operations was motivated by a desire to simplify the coefficient matrix in a manner analogous to 'eliminating unknowns' in the system of linear equations. Let us now make a formal definition of the type of matrix at which we were attempting to arrive.

Definition. *An* m × n *matrix* R *is called* **row-reduced** *if:*

(a) *the first non-zero entry in each non-zero row of* R *is equal to* 1;

(b) *each column of* R *which contains the leading non-zero entry of some row has all its other entries* 0.

EXAMPLE 7. One example of a row-reduced matrix is the $n \times n$ (square) **identity matrix** I. This is the $n \times n$ matrix defined by

$$I_{ij} = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

This is the first of many occasions on which we shall use the **Kronecker delta** (δ).

In Examples 5 and 6, the final matrices in the sequences exhibited there are row-reduced matrices. Two examples of matrices which are *not* row-reduced are:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The second matrix fails to satisfy condition (a), because the leading non-zero entry of the first row is not 1. The first matrix does satisfy condition (a), but fails to satisfy condition (b) in column 3.

We shall now prove that we can pass from any given matrix to a row-reduced matrix, by means of a finite number of elementary row operations. In combination with Theorem 3, this will provide us with an effective tool for solving systems of linear equations.

Theorem 4. *Every* m × n *matrix over the field* F *is row-equivalent to a row-reduced matrix.*

Proof. Let A be an $m \times n$ matrix over F. If every entry in the first row of A is 0, then condition (a) is satisfied in so far as row 1 is concerned. If row 1 has a non-zero entry, let k be the smallest positive integer j for which $A_{1j} \neq 0$. Multiply row 1 by A_{1k}^{-1}, and then condition (a) is satisfied with regard to row 1. Now for each $i \geq 2$, add $(-A_{ik})$ times row 1 to row i. Now the leading non-zero entry of row 1 occurs in column k, that entry is 1, and every other entry in column k is 0.

Now consider the matrix which has resulted from above. If every entry in row 2 is 0, we do nothing to row 2. If some entry in row 2 is different from 0, we multiply row 2 by a scalar so that the leading non-zero entry is 1. In the event that row 1 had a leading non-zero entry in column k, this leading non-zero entry of row 2 cannot occur in column k; say it occurs in column $k_r \neq k$. By adding suitable multiples of row 2 to the various rows, we can arrange that all entries in column k' are 0, except the 1 in row 2. The important thing to notice is this: In carrying out these last operations, we will not change the entries of row 1 in columns $1, \ldots, k$; nor will we change any entry of column k. Of course, if row 1 was identically 0, the operations with row 2 will not affect row 1.

Working with one row at a time in the above manner, it is clear that in a finite number of steps we will arrive at a row-reduced matrix. ∎

Exercises

1. Find all solutions to the system of equations

$$(1 - i)x_1 - ix_2 = 0$$
$$2x_1 + (1 - i)x_2 = 0.$$

2. If

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}$$

find all solutions of $AX = 0$ by row-reducing A.

3. If

$$A = \begin{bmatrix} 6 & -4 & 0 \\ 4 & -2 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

find all solutions of $AX = 2X$ and all solutions of $AX = 3X$. (The symbol cX denotes the matrix each entry of which is c times the corresponding entry of X.)

4. Find a row-reduced matrix which is row-equivalent to

$$A = \begin{bmatrix} i & -(1+i) & 0 \\ 1 & -2 & 1 \\ 1 & 2i & -1 \end{bmatrix}.$$

5. Prove that the following two matrices are *not* row-equivalent:

$$\begin{bmatrix} 2 & 0 & 0 \\ a & -1 & 0 \\ b & c & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 2 \\ -2 & 0 & -1 \\ 1 & 3 & 5 \end{bmatrix}.$$

6. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a 2×2 matrix with complex entries. Suppose that A is row-reduced and also that $a + b + c + d = 0$. Prove that there are exactly three such matrices.

7. Prove that the interchange of two rows of a matrix can be accomplished by a finite sequence of elementary row operations of the other two types.

8. Consider the system of equations $AX = 0$ where

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is a 2×2 matrix over the field F. Prove the following.

(a) If every entry of A is 0, then every pair (x_1, x_2) is a solution of $AX = 0$.

(b) If $ad - bc \neq 0$, the system $AX = 0$ has only the trivial solution $x_1 = x_2 = 0$.

(c) If $ad - bc = 0$ and some entry of A is different from 0, then there is a solution (x_1^0, x_2^0) such that (x_1, x_2) is a solution if and only if there is a scalar y such that $x_1 = yx_1^0$, $x_2 = yx_2^0$.

1.4. Row-Reduced Echelon Matrices

Until now, our work with systems of linear equations was motivated by an attempt to find the solutions of such a system. In Section 1.3 we established a standardized technique for finding these solutions. We wish now to acquire some information which is slightly more theoretical, and for that purpose it is convenient to go a little beyond row-reduced matrices.

Definition. *An* m × n *matrix* R *is called a* **row-reduced echelon matrix** *if:*

(a) R *is row-reduced;*

(b) *every row of* R *which has all its entries* 0 *occurs below every row which has a non-zero entry;*

(c) *if rows* 1, . . . , r *are the non-zero rows of* R, *and if the leading non-zero entry of row* i *occurs in column* k_i, i = 1, . . . , r, *then* $k_1 < k_2 < \cdots < k_r$.

One can also describe an $m \times n$ row-reduced echelon matrix R as follows. Either every entry in R is 0, or there exists a positive integer r, $1 \leq r \leq m$, and r positive integers $k_1, \ldots, k_r$ with $1 \leq k_i \leq n$ and

(a) $R_{ij} = 0$ for $i > r$, and $R_{ij} = 0$ if $j < k_i$.

(b) $R_{ik_i} = \delta_{ij}$, $1 \leq i \leq r$, $1 \leq j \leq r$.

(c) $k_1 < \cdots < k_r$.

EXAMPLE 8. Two examples of row-reduced echelon matrices are the $n \times n$ identity matrix, and the $m \times n$ **zero matrix** $0^{m,n}$, in which all entries are 0. The reader should have no difficulty in making other examples, but we should like to give one non-trivial one:

$$\begin{bmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Theorem 5. *Every* m × n *matrix* A *is row-equivalent to a row-reduced echelon matrix.*

Proof. We know that A is row-equivalent to a row-reduced matrix. All that we need observe is that by performing a finite number of row interchanges on a row-reduced matrix we can bring it to row-reduced echelon form. ∎

In Examples 5 and 6, we saw the significance of row-reduced matrices in solving homogeneous systems of linear equations. Let us now discuss briefly the system $RX = 0$, when R is a row-reduced echelon matrix. Let rows 1, . . . , r be the non-zero rows of R, and suppose that the leading non-zero entry of row i occurs in column k_i. The system $RX = 0$ then consists of r non-trivial equations. Also the unknown x_{k_i} will occur (with non-zero coefficient) only in the ith equation. If we let $u_1, \ldots, u_{n-r}$ denote the $(n - r)$ unknowns which are different from $x_{k_1}, \ldots, x_{k_r}$, then the r non-trivial equations in $RX = 0$ are of the form

$$\begin{aligned} x_{k_1} + \sum_{j=1}^{n-r} C_{1j}u_j &= 0 \\ &\vdots \\ x_{k_r} + \sum_{j=1}^{n-r} C_{rj}u_j &= 0. \end{aligned} \tag{1-3}$$

All the solutions to the system of equations $RX = 0$ are obtained by assigning any values whatsoever to $u_1, \ldots, u_{n-r}$ and then computing the corresponding values of $x_{k_1}, \ldots, x_{k_r}$ from (1-3). For example, if R is the matrix displayed in Example 8, then $r = 2$, $k_1 = 2$, $k_2 = 4$, and the two non-trivial equations in the system $RX = 0$ are

$$\begin{aligned} x_2 - 3x_3 \quad + \tfrac{1}{2}x_5 = 0 \quad &\text{or} \quad x_2 = 3x_3 - \tfrac{1}{2}x_5 \\ x_4 + 2x_5 = 0 \quad &\text{or} \quad x_4 = -2x_5. \end{aligned}$$

So we may assign any values to x_1, x_3, and x_5, say $x_1 = a$, $x_3 = b$, $x_5 = c$, and obtain the solution $(a, 3b - \frac{1}{2}c, b, -2c, c)$.

Let us observe one thing more in connection with the system of equations $RX = 0$. If the number r of non-zero rows in R is less than n, then the system $RX = 0$ has a non-trivial solution, that is, a solution $(x_1, \ldots, x_n)$ in which not every x_j is 0. For, since $r < n$, we can choose some x_j which is not among the r unknowns $x_{k_1}, \ldots, x_{k_r}$, and we can then construct a solution as above in which this x_j is 1. This observation leads us to one of the most fundamental facts concerning systems of homogeneous linear equations.

Theorem 6. *If* A *is an* m × n *matrix and* m < n, *then the homogeneous system of linear equations* AX = 0 *has a non-trivial solution.*

Proof. Let R be a row-reduced echelon matrix which is row-equivalent to A. Then the systems $AX = 0$ and $RX = 0$ have the same solutions by Theorem 3. If r is the number of non-zero rows in R, then certainly $r \leq m$, and since $m < n$, we have $r < n$. It follows immediately from our remarks above that $AX = 0$ has a non-trivial solution. ∎

Theorem 7. *If* A *is an* n × n *(square) matrix, then* A *is row-equivalent to the* n × n *identity matrix if and only if the system of equations* AX = 0 *has only the trivial solution.*

Proof. If A is row-equivalent to I, then $AX = 0$ and $IX = 0$ have the same solutions. Conversely, suppose $AX = 0$ has only the trivial solution $X = 0$. Let R be an $n \times n$ row-reduced echelon matrix which is row-equivalent to A, and let r be the number of non-zero rows of R. Then $RX = 0$ has no non-trivial solution. Thus $r \geq n$. But since R has n rows, certainly $r \leq n$, and we have $r = n$. Since this means that R actually has a leading non-zero entry of 1 in each of its n rows, and since these 1's occur each in a different one of the n columns, R must be the $n \times n$ identity matrix. ∎

Let us now ask what elementary row operations do toward solving a system of linear equations $AX = Y$ which is not homogeneous. At the outset, one must observe one basic difference between this and the homogeneous case, namely, that while the homogeneous system always has the

trivial solution $x_1 = \cdots = x_n = 0$, an inhomogeneous system need have no solution at all.

We form the **augmented matrix** A' of the system $AX = Y$. This is the $m \times (n + 1)$ matrix whose first n columns are the columns of A and whose last column is Y. More precisely,

$$A'_{ij} = A_{ij}, \quad \text{if} \quad j \leq n$$
$$A'_{i(n+1)} = y_i.$$

Suppose we perform a sequence of elementary row operations on A, arriving at a row-reduced echelon matrix R. If we perform this same sequence of row operations on the augmented matrix A', we will arrive at a matrix R' whose first n columns are the columns of R and whose last column contains certain scalars $z_1, \ldots, z_m$. The scalars z_i are the entries of the $m \times 1$ matrix

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

which results from applying the sequence of row operations to the matrix Y. It should be clear to the reader that, just as in the proof of Theorem 3, the systems $AX = Y$ and $RX = Z$ are equivalent and hence have the same solutions. It is very easy to determine whether the system $RX = Z$ has any solutions and to determine all the solutions if any exist. For, if R has r non-zero rows, with the leading non-zero entry of row i occurring in column k_i, $i = 1, \ldots, r$, then the first r equations of $RX = Z$ effectively express $x_{k_1}, \ldots, x_{k_r}$ in terms of the $(n - r)$ remaining x_j and the scalars $z_1, \ldots, z_r$. The last $(m - r)$ equations are

$$\begin{aligned} 0 &= z_{r+1} \\ &\vdots \\ 0 &= z_m \end{aligned}$$

and accordingly the condition for the system to have a solution is $z_i = 0$ for $i > r$. If this condition is satisfied, all solutions to the system are found just as in the homogeneous case, by assigning arbitrary values to $(n - r)$ of the x_j and then computing x_{k_i} from the ith equation.

EXAMPLE 9. Let F be the field of rational numbers and

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{bmatrix}$$

and suppose that we wish to solve the system $AX = Y$ for some y_1, y_2, and y_3. Let us perform a sequence of row operations on the augmented matrix A' which row-reduces A:

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 2 & 1 & 1 & y_2 \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix}.$$

The condition that the system $AX = Y$ have a solution is thus

$$2y_1 - y_2 + y_3 = 0$$

and if the given scalars y_i satisfy this condition, all solutions are obtained by assigning a value c to x_3 and then computing

$$\begin{aligned} x_1 &= -\tfrac{3}{5}c + \tfrac{1}{5}(y_1 + 2y_2) \\ x_2 &= \tfrac{1}{5}c + \tfrac{1}{5}(y_2 - 2y_1). \end{aligned}$$

Let us observe one final thing about the system $AX = Y$. Suppose the entries of the matrix A and the scalars $y_1, \ldots, y_m$ happen to lie in a subfield F_1 of the field F. If the system of equations $AX = Y$ has a solution with $x_1, \ldots, x_n$ in F, it has a solution with $x_1, \ldots, x_n$ in F_1. For, over either field, the condition for the system to have a solution is that certain relations hold between $y_1, \ldots, y_m$ in F_1 (the relations $z_i = 0$ for $i > r$, above). For example, if $AX = Y$ is a system of linear equations in which the scalars y_k and A_{ij} are real numbers, and if there is a solution in which $x_1, \ldots, x_n$ are complex numbers, then there is a solution with $x_1, \ldots, x_n$ real numbers.

Exercises

1. Find all solutions to the following system of equations by row-reducing the coefficient matrix:

$$\begin{aligned} \tfrac{1}{3}x_1 + 2x_2 - 6x_3 &= 0 \\ -4x_1 + 5x_3 &= 0 \\ -3x_1 + 6x_2 - 13x_3 &= 0 \\ -\tfrac{7}{3}x_1 + 2x_2 - \tfrac{8}{3}x_3 &= 0 \end{aligned}$$

2. Find a row-reduced echelon matrix which is row-equivalent to

$$A = \begin{bmatrix} 1 & -i \\ 2 & 2 \\ i & 1+i \end{bmatrix}.$$

What are the solutions of $AX = 0$?

3. Describe explicitly all 2×2 row-reduced echelon matrices.

4. Consider the system of equations

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 + 2x_3 &= 1 \\ x_1 - 3x_2 + 4x_3 &= 2. \end{aligned}$$

Does this system have a solution? If so, describe explicitly all solutions.

5. Give an example of a system of two linear equations in two unknowns which has no solution.

6. Show that the system

$$\begin{aligned} x_1 - 2x_2 + x_3 + 2x_4 &= 1 \\ x_1 + x_2 - x_3 + x_4 &= 2 \\ x_1 + 7x_2 - 5x_3 - x_4 &= 3 \end{aligned}$$

has no solution.

7. Find all solutions of

$$\begin{aligned} 2x_1 - 3x_2 - 7x_3 + 5x_4 + 2x_5 &= -2 \\ x_1 - 2x_2 - 4x_3 + 3x_4 + x_5 &= -2 \\ 2x_1 - 4x_3 + 2x_4 + x_5 &= 3 \\ x_1 - 5x_2 - 7x_3 + 6x_4 + 2x_5 &= -7. \end{aligned}$$

8. Let

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}.$$

For which triples (y_1, y_2, y_3) does the system $AX = Y$ have a solution?

9. Let

$$A = \begin{bmatrix} 3 & -6 & 2 & -1 \\ -2 & 4 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & -2 & 1 & 0 \end{bmatrix}.$$

For which (y_1, y_2, y_3, y_4) does the system of equations $AX = Y$ have a solution?

10. Suppose R and R' are 2×3 row-reduced echelon matrices and that the systems $RX = 0$ and $R'X = 0$ have exactly the same solutions. Prove that $R = R'$.

1.5. Matrix Multiplication

It is apparent (or should be, at any rate) that the process of forming linear combinations of the rows of a matrix is a fundamental one. For this reason it is advantageous to introduce a systematic scheme for indicating just what operations are to be performed. More specifically, suppose B is an $n \times p$ matrix over a field F with rows $\beta_1, \ldots, \beta_n$ and that from B we construct a matrix C with rows $\gamma_1, \ldots, \gamma_m$ by forming certain linear combinations

$$\gamma_i = A_{i1}\beta_1 + A_{i2}\beta_2 + \cdots + A_{in}\beta_n. \tag{1-4}$$

The rows of C are determined by the mn scalars A_{ij} which are themselves the entries of an $m \times n$ matrix A. If (1-4) is expanded to

$$(C_{i1} \cdots C_{ip}) = \sum_{r=1}^{n} (A_{ir}B_{r1} \cdots A_{ir}B_{rp})$$

we see that the entries of C are given by

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

Definition. *Let* A *be an* m × n *matrix over the field* F *and let* B *be an* n × p *matrix over* F. *The* **product** AB *is the* m × p *matrix* C *whose* i, j *entry is*

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

EXAMPLE 10. Here are some products of matrices with rational entries.

(a)
$$\begin{bmatrix} 5 & -1 & 2 \\ 0 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 5 & -1 & 2 \\ 15 & 4 & 8 \end{bmatrix}$$

Here

$$\begin{aligned} \gamma_1 &= (5 \quad -1 \quad 2) = 1 \cdot (5 \quad -1 \quad 2) + 0 \cdot (15 \quad 4 \quad 8) \\ \gamma_2 &= (0 \quad 7 \quad 2) = -3(5 \quad -1 \quad 2) + 1 \cdot (15 \quad 4 \quad 8) \end{aligned}$$

(b)
$$\begin{bmatrix} 0 & 6 & 1 \\ 9 & 12 & -8 \\ 12 & 62 & -3 \\ 3 & 8 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 3 \\ 5 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 1 \\ 3 & 8 & -2 \end{bmatrix}$$

Here

$$\begin{aligned} \gamma_2 &= (\ 9 \quad 12 \quad -8) = -2(0 \quad 6 \quad 1) + 3(3 \quad 8 \quad -2) \\ \gamma_3 &= (12 \quad 62 \quad -3) = \quad 5(0 \quad 6 \quad 1) + 4(3 \quad 8 \quad -2) \end{aligned}$$

(c)
$$\begin{bmatrix} 8 \\ 29 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

(d)
$$\begin{bmatrix} -2 & -4 \\ 6 & 12 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} [2 \quad 4]$$

Here

$$\gamma_2 = (6 \quad 12) = 3(2 \quad 4)$$

(e)
$$[2 \quad 4] \begin{bmatrix} -1 \\ 3 \end{bmatrix} = [10]$$

(f)
$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(g)
$$\begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 9 & 0 \end{bmatrix}$$

It is important to observe that the product of two matrices need not be defined; the product is defined if and only if the number of columns in the first matrix coincides with the number of rows in the second matrix. Thus it is meaningless to interchange the order of the factors in (a), (b), and (c) above. Frequently we shall write products such as AB without explicitly mentioning the sizes of the factors and in such cases it will be understood that the product is defined. From (d), (e), (f), (g) we find that even when the products AB and BA are both defined it need not be true that $AB = BA$; in other words, matrix multiplication is *not commutative.*

EXAMPLE 11.

(a) If I is the $m \times m$ identity matrix and A is an $m \times n$ matrix, $IA = A$.

(b) If I is the $n \times n$ identity matrix and A is an $m \times n$ matrix, $AI = A$.

(c) If $0^{k,m}$ is the $k \times m$ zero matrix, $0^{k,n} = 0^{k,m}A$. Similarly, $A0^{n,p} = 0^{m,p}$.

EXAMPLE 12. Let A be an $m \times n$ matrix over F. Our earlier shorthand notation, $AX = Y$, for systems of linear equations is consistent with our definition of matrix products. For if

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

with x_i in F, then AX is the $m \times 1$ matrix

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

such that $y_i = A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n$.

The use of column matrices suggests a notation which is frequently useful. If B is an $n \times p$ matrix, the columns of B are the $1 \times n$ matrices $B_1, \ldots, B_p$ defined by

$$B_j = \begin{bmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{bmatrix}, \qquad 1 \leq j \leq p.$$

The matrix B is the succession of these columns:

$$B = [B_1, \ldots, B_p].$$

The i, j entry of the product matrix AB is formed from the ith row of A

and the jth column of B. The reader should verify that the jth column of AB is AB_j:

$$AB = [AB_1, \ldots, AB_p].$$

In spite of the fact that a product of matrices depends upon the order in which the factors are written, it is independent of the way in which they are associated, as the next theorem shows.

Theorem 8. *If* A, B, C *are matrices over the field* F *such that the products* BC *and* A(BC) *are defined, then so are the products* AB, (AB)C *and*

$$\mathrm{A(BC)} = \mathrm{(AB)C}.$$

Proof. Suppose B is an $n \times p$ matrix. Since BC is defined, C is a matrix with p rows, and BC has n rows. Because $A(BC)$ is defined we may assume A is an $m \times n$ matrix. Thus the product AB exists and is an $m \times p$ matrix, from which it follows that the product $(AB)C$ exists. To show that $A(BC) = (AB)C$ means to show that

$$[A(BC)]_{ij} = [(AB)C]_{ij}$$

for each i, j. By definition

$$\begin{aligned}
[A(BC)]_{ij} &= \sum_r A_{ir}(BC)_{rj} \\
&= \sum_r A_{ir} \sum_s B_{rs}C_{sj} \\
&= \sum_r \sum_s A_{ir}B_{rs}C_{sj} \\
&= \sum_s \sum_r A_{ir}B_{rs}C_{sj} \\
&= \sum_s \Big(\sum_r A_{ir}B_{rs}\Big)C_{sj} \\
&= \sum_s (AB)_{is}C_{sj} \\
&= [(AB)C]_{ij}. \quad \blacksquare
\end{aligned}$$

When A is an $n \times n$ (square) matrix, the product AA is defined. We shall denote this matrix by A^2. By Theorem 8, $(AA)A = A(AA)$ or $A^2A = AA^2$, so that the product AAA is unambiguously defined. This product we denote by A^3. In general, the product $AA \cdots A$ (k times) is unambiguously defined, and we shall denote this product by A^k.

Note that the relation $A(BC) = (AB)C$ implies among other things that linear combinations of linear combinations of the rows of C are again linear combinations of the rows of C.

If B is a given matrix and C is obtained from B by means of an elementary row operation, then each row of C is a linear combination of the rows of B, and hence there is a matrix A such that $AB = C$. In general there are many such matrices A, and among all such it is convenient and

possible to choose one having a number of special properties. Before going into this we need to introduce a class of matrices.

Definition. *An* m × n *matrix is said to be an* **elementary matrix** *if it can be obtained from the* m × m *identity matrix by means of a single elementary row operation.*

EXAMPLE 13. A 2×2 elementary matrix is necessarily one of the following:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \quad c \neq 0, \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}, \quad c \neq 0.$$

Theorem 9. *Let* e *be an elementary row operation and let* E *be the* m × m *elementary matrix* E = e(I). *Then, for every* m × n *matrix* A,

$$e(A) = EA.$$

Proof. The point of the proof is that the entry in the ith row and jth column of the product matrix EA is obtained from the ith row of E and the jth column of A. The three types of elementary row operations should be taken up separately. We shall give a detailed proof for an operation of type (ii). The other two cases are even easier to handle than this one and will be left as exercises. Suppose $r \neq s$ and e is the operation 'replacement of row r by row r plus c times row s.' Then

$$E_{ik} = \begin{cases} \delta_{ik}, & i \neq r \\ \delta_{rk} + c\delta_{sk}, & i = r. \end{cases}$$

Therefore,

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} A_{ik}, & i \neq r \\ A_{rj} + cA_{sj}, & i = r. \end{cases}$$

In other words $EA = e(A)$. ∎

Corollary. *Let* A *and* B *be* m × n *matrices over the field* F. *Then* B *is row-equivalent to* A *if and only if* B = PA, *where* P *is a product of* m × m *elementary matrices.*

Proof. Suppose $B = PA$ where $P = E_s \cdots E_2E_1$ and the E_i are $m \times m$ elementary matrices. Then E_1A is row-equivalent to A, and $E_2(E_1A)$ is row-equivalent to E_1A. So E_2E_1A is row-equivalent to A; and continuing in this way we see that $(E_s \cdots E_1)A$ is row-equivalent to A.

Now suppose that B is row-equivalent to A. Let $E_1, E_2, \ldots, E_s$ be the elementary matrices corresponding to some sequence of elementary row operations which carries A into B. Then $B = (E_s \cdots E_1)A$. ∎

Exercises

1. Let

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}, \qquad C = [1 \quad -1].$$

Compute ABC and CAB.

2. Let

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & -2 \\ 1 & 3 \\ 4 & 4 \end{bmatrix}.$$

Verify directly that $A(AB) = A^2B$.

3. Find two different 2×2 matrices A such that $A^2 = 0$ but $A \neq 0$.

4. For the matrix A of Exercise 2, find elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_k \cdots E_2E_1A = I.$$

5. Let

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 2 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 & 1 \\ -4 & 4 \end{bmatrix}.$$

Is there a matrix C such that $CA = B$?

6. Let A be an $m \times n$ matrix and B an $n \times k$ matrix. Show that the columns of $C = AB$ are linear combinations of the columns of A. If $\alpha_1, \ldots, \alpha_n$ are the columns of A and $\gamma_1, \ldots, \gamma_k$ are the columns of C, then

$$\gamma_j = \sum_{r=1}^{n} B_{rj}\alpha_r.$$

7. Let A and B be 2×2 matrices such that $AB = I$. Prove that $BA = I$.

8. Let

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

be a 2×2 matrix. We inquire when it is possible to find 2×2 matrices A and B such that $C = AB - BA$. Prove that such matrices can be found if and only if $C_{11} + C_{22} = 0$.

1.6. Invertible Matrices

Suppose P is an $m \times m$ matrix which is a product of elementary matrices. For each $m \times n$ matrix A, the matrix $B = PA$ is row-equivalent to A; hence A is row-equivalent to B and there is a product Q of elementary matrices such that $A = QB$. In particular this is true when A is the

$m \times m$ identity matrix. In other words, there is an $m \times m$ matrix Q, which is itself a product of elementary matrices. such that $QP = I$. As we shall soon see, the existence of a Q with $QP = I$ is equivalent to the fact that P is a product of elementary matrices.

Definition. *Let* A *be an* n $\times$ n *(square) matrix over the field* F. *An* n $\times$ n *matrix* B *such that* BA = I *is called a* **left inverse** *of* A; *an* n $\times$ n *matrix* B *such that* AB = I *is called a* **right inverse** *of* A. *If* AB = BA = I, *then* B *is called a* **two-sided inverse** *of* A *and* A *is said to be* **invertible.**

Lemma. *If* A *has a left inverse* B *and a right inverse* C, *then* B = C.

Proof. Suppose $BA = I$ and $AC = I$. Then

$$B = BI = B(AC) = (BA)C = IC = C. \quad \blacksquare$$

Thus if A has a left and a right inverse, A is invertible and has a unique two-sided inverse, which we shall denote by A^{-1} and simply call **the inverse** of A.

Theorem 10. *Let* A *and* B *be* n $\times$ n *matrices over* F.

(i) *If* A *is invertible, so is* A^{-1} *and* $(A^{-1})^{-1} = A$.

(ii) *If both* A *and* B *are invertible, so is* AB, *and* $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. The first statement is evident from the symmetry of the definition. The second follows upon verification of the relations

$$(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I. \quad \blacksquare$$

Corollary. *A product of invertible matrices is invertible.*

Theorem 11. *An elementary matrix is invertible.*

Proof. Let E be an elementary matrix corresponding to the elementary row operation e. If e_1 is the inverse operation of e (Theorem 2) and $E_1 = e_1(I)$, then

$$EE_1 = e(E_1) = e(e_1(I)) = I$$

and

$$E_1E = e_1(E) = e_1(e(I)) = I$$

so that E is invertible and $E_1 = E^{-1}$. $\blacksquare$

EXAMPLE 14.

(a)
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(b)
$$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -c \\ 0 & 1 \end{bmatrix}$$

(c) $$\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}$$

(d) When $c \neq 0$,

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} c^{-1} & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & c^{-1} \end{bmatrix}.$$

Theorem 12. *If* A *is an* n × n *matrix, the following are equivalent.*

(i) A *is invertible.*

(ii) A *is row-equivalent to the* n × n *identity matrix.*

(iii) A *is a product of elementary matrices.*

Proof. Let R be a row-reduced echelon matrix which is row-equivalent to A. By Theorem 9 (or its corollary),

$$R = E_k \cdots E_2 E_1 A$$

where $E_1, \ldots, E_k$ are elementary matrices. Each E_j is invertible, and so

$$A = E_1^{-1} \cdots E_k^{-1} R.$$

Since products of invertible matrices are invertible, we see that A is invertible if and only if R is invertible. Since R is a (square) row-reduced echelon matrix, R is invertible if and only if each row of R contains a non-zero entry, that is, if and only if $R = I$. We have now shown that A is invertible if and only if $R = I$, and if $R = I$ then $A = E_k^{-1} \cdots E_1^{-1}$. It should now be apparent that (i), (ii), and (iii) are equivalent statements about A. ∎

Corollary. *If* A *is an invertible* n × n *matrix and if a sequence of elementary row operations reduces* A *to the identity, then that same sequence of operations when applied to* I *yields* A^{-1}.

Corollary. *Let* A *and* B *be* m × n *matrices. Then* B *is row-equivalent to* A *if and only if* B = PA *where* P *is an invertible* m × m *matrix.*

Theorem 13. *For an* n × n *matrix* A, *the following are equivalent.*

(i) A *is invertible.*

(ii) *The homogeneous system* AX = 0 *has only the trivial solution* X = 0.

(iii) *The system of equations* AX = Y *has a solution* X *for each* n × 1 *matrix* Y.

Proof. According to Theorem 7, condition (ii) is equivalent to the fact that A is row-equivalent to the identity matrix. By Theorem 12, (i) and (ii) are therefore equivalent. If A is invertible, the solution of $AX = Y$ is $X = A^{-1}Y$. Conversely, suppose $AX = Y$ has a solution for each given Y. Let R be a row-reduced echelon matrix which is row-

equivalent to A. We wish to show that $R = I$. That amounts to showing that the last row of R is not (identically) 0. Let

$$E = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

If the system $RX = E$ can be solved for X, the last row of R cannot be 0. We know that $R = PA$, where P is invertible. Thus $RX = E$ if and only if $AX = P^{-1}E$. According to (iii), the latter system has a solution. ∎

Corollary. *A square matrix with either a left or right inverse is invertible.*

Proof. Let A be an $n \times n$ matrix. Suppose A has a left inverse, i.e., a matrix B such that $BA = I$. Then $AX = 0$ has only the trivial solution, because $X = IX = B(AX)$. Therefore A is invertible. On the other hand, suppose A has a right inverse, i.e., a matrix C such that $AC = I$. Then C has a left inverse and is therefore invertible. It then follows that $A = C^{-1}$ and so A is invertible with inverse C. ∎

Corollary. *Let* $A = A_1A_2 \cdots A_k$, *where* $A_1 \ldots, A_k$ *are* n $\times$ n (*square*) *matrices. Then* A *is invertible if and only if each* A_j *is invertible.*

Proof. We have already shown that the product of two invertible matrices is invertible. From this one sees easily that if each A_j is invertible then A is invertible.

Suppose now that A is invertible. We first prove that A_k is invertible. Suppose X is an $n \times 1$ matrix and $A_kX = 0$. Then $AX = (A_1 \cdots A_{k-1})A_kX = 0$. Since A is invertible we must have $X = 0$. The system of equations $A_kX = 0$ thus has no non-trivial solution, so A_k is invertible. But now $A_1 \cdots A_{k-1} = AA_k^{-1}$ is invertible. By the preceding argument, A_{k-1} is invertible. Continuing in this way, we conclude that each A_j is invertible. ∎

We should like to make one final comment about the solution of linear equations. Suppose A is an $m \times n$ matrix and we wish to solve the system of equations $AX = Y$. If R is a row-reduced echelon matrix which is row-equivalent to A, then $R = PA$ where P is an $m \times m$ invertible matrix. The solutions of the system $AX = Y$ are exactly the same as the solutions of the system $RX = PY$ $(= Z)$. In practice, it is not much more difficult to find the matrix P than it is to row-reduce A to R. For, suppose we form the augmented matrix A' of the system $AX = Y$, with arbitrary scalars $y_1, \ldots, y_m$ occurring in the last column. If we then perform on A' a sequence of elementary row operations which leads from A to R, it will

become evident what the matrix P is. (The reader should refer to Example 9 where we essentially carried out this process.) In particular, if A is a square matrix, this process will make it clear whether or not A is invertible and if A is invertible what the inverse P is. Since we have already given the nucleus of one example of such a computation, we shall content ourselves with a 2×2 example.

EXAMPLE 15. Suppose F is the field of rational numbers and

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}.$$

Then

$$\begin{bmatrix} 2 & -1 & y_1 \\ 1 & 3 & y_2 \end{bmatrix} \xrightarrow{(3)} \begin{bmatrix} 1 & 3 & y_2 \\ 2 & -1 & y_1 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 3 & y_2 \\ 0 & -7 & y_1 - 2y_2 \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 1 & 3 & y_2 \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 0 & \frac{1}{7}(y_2 + 3y_1) \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix}$$

from which it is clear that A is invertible and

$$A^{-1} = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{2}{7} \end{bmatrix}.$$

It may seem cumbersome to continue writing the arbitrary scalars $y_1, y_2, \ldots$ in the computation of inverses. Some people find it less awkward to carry along two sequences of matrices, one describing the reduction of A to the identity and the other recording the effect of the same sequence of operations starting from the identity. The reader may judge for himself which is a neater form of bookkeeping.

EXAMPLE 16. Let us find the inverse of

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{12} & \frac{4}{45} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & 0 & \frac{1}{180} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ \frac{1}{6} & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -6 & 12 & 0 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -9 & 60 & -60 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

It must have occurred to the reader that we have carried on a lengthy discussion of the rows of matrices and have said little about the columns. We focused our attention on the rows because this seemed more natural from the point of view of linear equations. Since there is obviously nothing sacred about rows, the discussion in the last sections could have been carried on using columns rather than rows. If one defines an elementary column operation and column-equivalence in a manner analogous to that of elementary row operation and row-equivalence, it is clear that each $m \times n$ matrix will be column-equivalent to a 'column-reduced echelon' matrix. Also each elementary column operation will be of the form $A \to AE$, where E is an $n \times n$ elementary matrix—and so on.

Exercises

1. Let

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ -1 & 0 & 3 & 5 \\ 1 & -2 & 1 & 1 \end{bmatrix}.$$

Find a row-reduced echelon matrix R which is row-equivalent to A and an invertible 3×3 matrix P such that $R = PA$.

2. Do Exercise 1, but with

$$A = \begin{bmatrix} 2 & 0 & i \\ 1 & -3 & -i \\ i & 1 & 1 \end{bmatrix}.$$

3. For each of the two matrices

$$\begin{bmatrix} 2 & 5 & -1 \\ 4 & -1 & 2 \\ 6 & 4 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & 4 \\ 0 & 1 & -2 \end{bmatrix}$$

use elementary row operations to discover whether it is invertible, and to find the inverse in case it is.

4. Let

$$A = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}.$$

For which X does there exist a scalar c such that $AX = cX$?

5. Discover whether

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

is invertible, and find A^{-1} if it exists.

6. Suppose A is a 2×1 matrix and that B is a 1×2 matrix. Prove that $C = AB$ is not invertible.

7. Let A be an $n \times n$ (square) matrix. Prove the following two statements:
(a) If A is invertible and $AB = 0$ for some $n \times n$ matrix B, then $B = 0$.
(b) If A is not invertible, then there exists an $n \times n$ matrix B such that $AB = 0$ but $B \neq 0$.

8. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Prove, using elementary row operations, that A is invertible if and only if $(ad - bc) \neq 0$.

9. An $n \times n$ matrix A is called **upper-triangular** if $A_{ij} = 0$ for $i > j$, that is, if every entry below the main diagonal is 0. Prove that an upper-triangular (square) matrix is invertible if and only if every entry on its main diagonal is different from 0.

10. Prove the following generalization of Exercise 6. If A is an $m \times n$ matrix, B is an $n \times m$ matrix and $n < m$, then AB is not invertible.

11. Let A be an $m \times n$ matrix. Show that by means of a finite number of elementary row and/or column operations one can pass from A to a matrix R which is both 'row-reduced echelon' and 'column-reduced echelon,' i.e., $R_{ij} = 0$ if $i \neq j$, $R_{ii} = 1$, $1 \leq i \leq r$, $R_{ii} = 0$ if $i > r$. Show that $R = PAQ$, where P is an invertible $m \times m$ matrix and Q is an invertible $n \times n$ matrix.

12. The result of Example 16 suggests that perhaps the matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

is invertible and A^{-1} has integer entries. Can you prove that?

https://www.facebook.com/GrinchaZ

2. Vector Spaces

2.1. Vector Spaces

In various parts of mathematics, one is confronted with a set, such that it is both meaningful and interesting to deal with 'linear combinations' of the objects in that set. For example, in our study of linear equations we found it quite natural to consider linear combinations of the rows of a matrix. It is likely that the reader has studied calculus and has dealt there with linear combinations of functions; certainly this is so if he has studied differential equations. Perhaps the reader has had some experience with vectors in three-dimensional Euclidean space, and in particular, with linear combinations of such vectors.

Loosely speaking, linear algebra is that branch of mathematics which treats the common properties of algebraic systems which consist of a set, together with a reasonable notion of a 'linear combination' of elements in the set. In this section we shall define the mathematical object which experience has shown to be the most useful abstraction of this type of algebraic system.

Definition. *A* **vector space** *(or linear space) consists of the following:*

1. *a field* F *of scalars;*
2. *a set* V *of objects, called vectors;*
3. *a rule (or operation), called vector addition, which associates with each pair of vectors* α, β *in* V *a vector* $\alpha + \beta$ *in* V, *called the sum of* α *and* β, *in such a way that*

   (a) *addition is commutative,* $\alpha + \beta = \beta + \alpha$;

   (b) *addition is associative,* $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$;

(c) *there is a unique vector* 0 *in* V, *called the zero vector, such that* $\alpha + 0 = \alpha$ *for all* α *in* V;

(d) *for each vector* α *in* V *there is a unique vector* $-\alpha$ *in* V *such that* $\alpha + (-\alpha) = 0$;

4. *a rule (or operation), called scalar multiplication, which associates with each scalar* c *in* F *and vector* α *in* V *a vector* $c\alpha$ *in* V, *called the product of* c *and* α, *in such a way that*

(a) $1\alpha = \alpha$ *for every* α *in* V;

(b) $(c_1c_2)\alpha = c_1(c_2\alpha)$;

(c) $c(\alpha + \beta) = c\alpha + c\beta$;

(d) $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$.

It is important to observe, as the definition states, that a vector space is a composite object consisting of a field, a set of 'vectors,' and two operations with certain special properties. The same set of vectors may be part of a number of distinct vector spaces (see Example 5 below). When there is no chance of confusion, we may simply refer to the vector space as V, or when it is desirable to specify the field, we shall say V is a **vector space over the field** F. The name 'vector' is applied to the elements of the set V largely as a matter of convenience. The origin of the name is to be found in Example 1 below, but one should not attach too much significance to the name, since the variety of objects occurring as the vectors in V may not bear much resemblance to any preassigned concept of vector which the reader has. We shall try to indicate this variety by a list of examples; our list will be enlarged considerably as we begin to study vector spaces.

EXAMPLE 1. **The n-tuple space, F^n.** Let F be any field, and let V be the set of all n-tuples $\alpha = (x_1, x_2, \ldots, x_n)$ of scalars x_i in F. If $\beta = (y_1, y_2, \ldots, y_n)$ with y_i in F, the sum of α and β is defined by

$$\alpha + \beta = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{2-1}$$

The product of a scalar c and vector α is defined by

$$c\alpha = (cx_1, cx_2, \ldots, cx_n). \tag{2-2}$$

The fact that this vector addition and scalar multiplication satisfy conditions (3) and (4) is easy to verify, using the similar properties of addition and multiplication of elements of F.

EXAMPLE 2. **The space of $m \times n$ matrices, $F^{m\times n}$.** Let F be any field and let m and n be positive integers. Let $F^{m\times n}$ be the set of all $m \times n$ matrices over the field F. The sum of two vectors A and B in $F^{m\times n}$ is defined by

$$(A + B)_{ij} = A_{ij} + B_{ij}. \tag{2-3}$$

The product of a scalar c and the matrix A is defined by

$$(cA)_{ij} = cA_{ij}. \tag{2-4}$$

Note that $F^{1\times n} = F^n$.

EXAMPLE 3. **The space of functions from a set to a field.** Let F be any field and let S be any non-empty set. Let V be the set of all functions from the set S into F. The sum of two vectors f and g in V is the vector $f + g$, i.e., the function from S into F, defined by

$$(f + g)(s) = f(s) + g(s). \tag{2-5}$$

The product of the scalar c and the function f is the function cf defined by

$$(cf)(s) = cf(s). \tag{2-6}$$

The preceding examples are special cases of this one. For an n-tuple of elements of F may be regarded as a function from the set S of integers $1, \ldots, n$ into F. Similarly, an $m \times n$ matrix over the field F is a function from the set S of pairs of integers, (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, into the field F. For this third example we shall indicate how one verifies that the operations we have defined satisfy conditions (3) and (4). For vector addition:

(a) Since addition in F is commutative,

$$f(s) + g(s) = g(s) + f(s)$$

for each s in S, so the functions $f + g$ and $g + f$ are identical.

(b) Since addition in F is associative,

$$f(s) + [g(s) + h(s)] = [f(s) + g(s)] + h(s)$$

for each s, so $f + (g + h)$ is the same function as $(f + g) + h$.

(c) The unique zero vector is the zero function which assigns to each element of S the scalar 0 in F.

(d) For each f in V, $(-f)$ is the function which is given by

$$(-f)(s) = -f(s).$$

The reader should find it easy to verify that scalar multiplication satisfies the conditions of (4), by arguing as we did with the vector addition.

EXAMPLE 4. **The space of polynomial functions over a field F.** Let F be a field and let V be the set of all functions f from F into F which have a rule of the form

$$f(x) = c_0 + c_1x + \cdots + c_nx^n \tag{2-7}$$

where $c_0, c_1, \ldots, c_n$ are fixed scalars in F (independent of x). A function of this type is called a **polynomial function on F**. Let addition and scalar multiplication be defined as in Example 3. One must observe here that if f and g are polynomial functions and c is in F, then $f + g$ and cf are again polynomial functions.

EXAMPLE 5. The field C of complex numbers may be regarded as a vector space over the field R of real numbers. More generally, let F be the field of real numbers and let V be the set of n-tuples $\alpha = (x_1, \ldots, x_n)$ where $x_1, \ldots, x_n$ are *complex* numbers. Define addition of vectors and scalar multiplication by (2-1) and (2-2), as in Example 1. In this way we obtain a vector space over the field R which is quite different from the space C^n and the space R^n.

There are a few simple facts which follow almost immediately from the definition of a vector space, and we proceed to derive these. If c is a scalar and 0 is the zero vector, then by 3(c) and 4(c)

$$c0 = c(0 + 0) = c0 + c0.$$

Adding $-(c0)$ and using 3(d), we obtain

$$c0 = 0. \tag{2-8}$$

Similarly, for the scalar 0 and any vector α we find that

$$0\alpha = 0. \tag{2-9}$$

If c is a non-zero scalar and α is a vector such that $c\alpha = 0$, then by (2-8), $c^{-1}(c\alpha) = 0$. But

$$c^{-1}(c\alpha) = (c^{-1}c)\alpha = 1\alpha = \alpha$$

hence, $\alpha = 0$. Thus we see that if c is a scalar and α a vector such that $c\alpha = 0$, then either c is the zero scalar or α is the zero vector.

If α is any vector in V, then

$$0 = 0\alpha = (1 - 1)\alpha = 1\alpha + (-1)\alpha = \alpha + (-1)\alpha$$

from which it follows that

$$(-1)\alpha = -\alpha. \tag{2-10}$$

Finally, the associative and commutative properties of vector addition imply that a sum involving a number of vectors is independent of the way in which these vectors are combined and associated. For example, if α_1, α_2, α_3, α_4 are vectors in V, then

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = [\alpha_2 + (\alpha_1 + \alpha_3)] + \alpha_4$$

and such a sum may be written without confusion as

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4.$$

Definition. *A vector β in* V *is said to be a* **linear combination** *of the vectors $\alpha_1, \ldots, \alpha_n$ in* V *provided there exist scalars* $c_1, \ldots, c_n$ *in* F *such that*

$$\beta = c_1\alpha_1 + \cdots + c_n\alpha_n$$

$$= \sum_{i=1}^{n} c_i\alpha_i.$$

Other extensions of the associative property of vector addition and the distributive properties 4(c) and 4(d) of scalar multiplication apply to linear combinations:

$$\sum_{i=1}^{n} c_i\alpha_i + \sum_{i=1}^{n} d_i\alpha_i = \sum_{i=1}^{n} (c_i + d_i)\alpha_i$$

$$c \sum_{i=1}^{n} c_i\alpha_i = \sum_{i=1}^{n} (cc_i)\alpha_i.$$

Certain parts of linear algebra are intimately related to geometry. The very word 'space' suggests something geometrical, as does the word 'vector' to most people. As we proceed with our study of vector spaces, the reader will observe that much of the terminology has a geometrical connotation. Before concluding this introductory section on vector spaces, we shall consider the relation of vector spaces to geometry to an extent which will at least indicate the origin of the name 'vector space.' This will be a brief intuitive discussion.

Let us consider the vector space R^3. In analytic geometry, one identifies triples (x_1, x_2, x_3) of real numbers with the points in three-dimensional Euclidean space. In that context, a vector is usually defined as a directed line segment PQ, from a point P in the space to another point Q. This amounts to a careful formulation of the idea of the 'arrow' from P to Q. As vectors are used, it is intended that they should be determined by their length and direction. Thus one must identify two directed line segments if they have the same length and the same direction.

The directed line segment PQ, from the point $P = (x_1, x_2, x_3)$ to the point $Q = (y_1, y_2, y_3)$, has the same length and direction as the directed line segment from the origin $O = (0, 0, 0)$ to the point $(y_1 - x_1, y_2 - x_2, y_3 - x_3)$. Furthermore, this is the only segment emanating from the origin which has the same length and direction as PQ. Thus, if one agrees to treat only vectors which emanate from the origin, there is exactly one vector associated with each given length and direction.

The vector OP, from the origin to $P = (x_1, x_2, x_3)$, is completely determined by P, and it is therefore possible to identify this vector with the point P. In our definition of the vector space R^3, the vectors are simply defined to be the triples (x_1, x_2, x_3).

Given points $P = (x_1, x_2, x_3)$ and $Q = (y_1, y_2, y_3)$, the definition of the sum of the vectors OP and OQ can be given geometrically. If the vectors are not parallel, then the segments OP and OQ determine a plane and these segments are two of the edges of a parallelogram in that plane (see Figure 1). One diagonal of this parallelogram extends from O to a point S, and the sum of OP and OQ is defined to be the vector OS. The coordinates of the point S are $(x_1 + y_1, x_2 + y_2, x_3 + y_3)$ and hence this geometrical definition of vector addition is equivalent to the algebraic definition of Example 1.

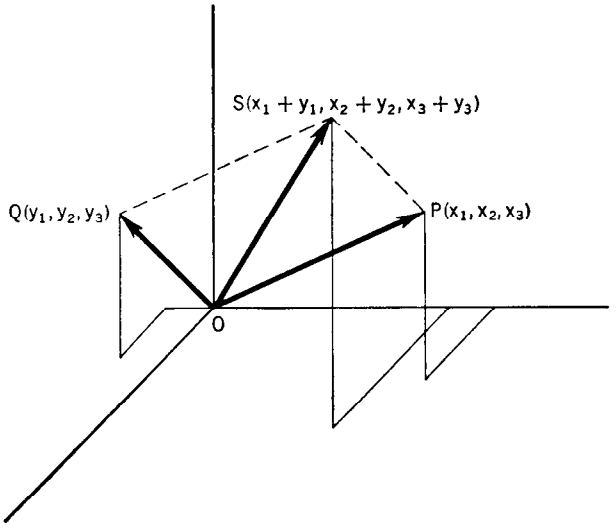

FIGURE 1

Scalar multiplication has a simpler geometric interpretation. If c is a real number, then the product of c and the vector OP is the vector from the origin with length $|c|$ times the length of OP and a direction which agrees with the direction of OP if $c > 0$, and which is opposite to the direction of OP if $c < 0$. This scalar multiplication just yields the vector OT where $T = (cx_1, cx_2, cx_3)$, and is therefore consistent with the algebraic definition given for R^3.

From time to time, the reader will probably find it helpful to 'think geometrically' about vector spaces, that is, to draw pictures for his own benefit to illustrate and motivate some of the ideas. Indeed, he should do this. However, in forming such illustrations he must bear in mind that, because we are dealing with vector spaces as algebraic systems, all proofs we give will be of an algebraic nature.

Exercises

1. If F is a field, verify that F^n (as defined in Example 1) is a vector space over the field F.

2. If V is a vector space over the field F, verify that

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = [\alpha_2 + (\alpha_3 + \alpha_1)] + \alpha_4$$

for all vectors α_1, α_2, α_3, and α_4 in V.

3. If C is the field of complex numbers, which vectors in C^3 are linear combinations of $(1, 0, -1)$, $(0, 1, 1)$, and $(1, 1, 1)$?

4. Let V be the set of all pairs (x, y) of real numbers, and let F be the field of real numbers. Define

$$(x, y) + (x_1, y_1) = (x + x_1, y + y_1)$$
$$c(x, y) = (cx, y).$$

Is V, with these operations, a vector space over the field of real numbers?

5. On R^n, define two operations

$$\alpha \oplus \beta = \alpha - \beta$$
$$c \cdot \alpha = -c\alpha.$$

The operations on the right are the usual ones. Which of the axioms for a vector space are satisfied by $(R^n, \oplus, \cdot)$?

6. Let V be the set of all complex-valued functions f on the real line such that (for all t in R)

$$f(-t) = \overline{f(t)}.$$

The bar denotes complex conjugation. Show that V, with the operations

$$(f + g)(t) = f(t) + g(t)$$
$$(cf)(t) = cf(t)$$

is a vector space over the field of *real* numbers. Give an example of a function in V which is not real-valued.

7. Let V be the set of pairs (x, y) of real numbers and let F be the field of real numbers. Define

$$(x, y) + (x_1, y_1) = (x + x_1, 0)$$
$$c(x, y) = (cx, 0).$$

Is V, with these operations, a vector space?

2.2. Subspaces

In this section we shall introduce some of the basic concepts in the study of vector spaces.

Definition. *Let* V *be a vector space over the field* F. *A* **subspace** *of* V *is a subset* W *of* V *which is itself a vector space over* F *with the operations of vector addition and scalar multiplication on* V.

A direct check of the axioms for a vector space shows that the subset W of V is a subspace if for each α and β in W the vector $\alpha + \beta$ is again in W; the 0 vector is in W; for each α in W the vector $(-\alpha)$ is in W; for each α in W and each scalar c the vector $c\alpha$ is in W. The commutativity and associativity of vector addition, and the properties (4)(a), (b), (c), and (d) of scalar multiplication do not need to be checked, since these are properties of the operations on V. One can simplify things still further.

Theorem 1. *A non-empty subset* W *of* V *is a subspace of* V *if and only if for each pair of vectors* α, β *in* W *and each scalar* c *in* F *the vector* $c\alpha + \beta$ *is again in* W.

Proof. Suppose that W is a non-empty subset of V such that $c\alpha + \beta$ belongs to W for all vectors α, β in W and all scalars c in F. Since W is non-empty, there is a vector ρ in W, and hence $(-1)\rho + \rho = 0$ is in W. Then if α is any vector in W and c any scalar, the vector $c\alpha = c\alpha + 0$ is in W. In particular, $(-1)\alpha = -\alpha$ is in W. Finally, if α and β are in W, then $\alpha + \beta = 1\alpha + \beta$ is in W. Thus W is a subspace of V.

Conversely, if W is a subspace of V, α and β are in W, and c is a scalar, certainly $c\alpha + \beta$ is in W. ∎

Some people prefer to use the $c\alpha + \beta$ property in Theorem 1 as the definition of a subspace. It makes little difference. The important point is that, if W is a non-empty subset of V such that $c\alpha + \beta$ is in V for all α, β in W and all c in F, then (with the operations inherited from V) W is a vector space. This provides us with many new examples of vector spaces.

EXAMPLE 6.

(a) If V is any vector space, V is a subspace of V; the subset consisting of the zero vector alone is a subspace of V, called the **zero subspace** of V.

(b) In F^n, the set of n-tuples $(x_1, \ldots, x_n)$ with $x_1 = 0$ is a subspace; however, the set of n-tuples with $x_1 = 1 + x_2$ is not a subspace ($n \geq 2$).

(c) The space of polynomial functions over the field F is a subspace of the space of all functions from F into F.

(d) An $n \times n$ (square) matrix A over the field F is **symmetric** if $A_{ij} = A_{ji}$ for each i and j. The symmetric matrices form a subspace of the space of all $n \times n$ matrices over F.

(e) An $n \times n$ (square) matrix A over the field C of complex numbers is **Hermitian** (or **self-adjoint**) if

$$A_{jk} = \overline{A_{kj}}$$

for each j, k, the bar denoting complex conjugation. A 2×2 matrix is Hermitian if and only if it has the form

$$\begin{bmatrix} z & x + iy \\ x - iy & w \end{bmatrix}$$

where x, y, z, and w are real numbers. The set of all Hermitian matrices is *not* a subspace of the space of all $n \times n$ matrices over C. For if A is Hermitian, its diagonal entries $A_{11}, A_{22}, \ldots$, are all real numbers, but the diagonal entries of iA are in general not real. On the other hand, it is easily verified that the set of $n \times n$ complex Hermitian matrices is a vector space over the field R of real numbers (with the usual operations).

EXAMPLE 7. **The solution space of a system of homogeneous linear equations.** Let A be an $m \times n$ matrix over F. Then the set of all $n \times 1$ (column) matrices X over F such that $AX = 0$ is a subspace of the space of all $n \times 1$ matrices over F. To prove this we must show that $A(cX + Y) = 0$ when $AX = 0$, $AY = 0$, and c is an arbitrary scalar in F. This follows immediately from the following general fact.

Lemma. *If* A *is an* m $\times$ n *matrix over* F *and* B, C *are* n $\times$ p *matrices over* F *then*

$$\text{A(dB + C)} = \text{d(AB)} + \text{AC} \tag{2-11}$$

for each scalar d *in* F.

Proof.

$$\begin{aligned}
[A(dB + C)]_{ij} &= \sum_k A_{ik}(dB + C)_{kj} \\
&= \sum_k (dA_{ik}B_{kj} + A_{ik}C_{kj}) \\
&= d\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\
&= d(AB)_{ij} + (AC)_{ij} \\
&= [d(AB) + AC]_{ij}. \quad \blacksquare
\end{aligned}$$

Similarly one can show that $(dB + C)A = d(BA) + CA$, if the matrix sums and products are defined.

Theorem 2. *Let* V *be a vector space over the field* F. *The intersection of any collection of subspaces of* V *is a subspace of* V.

Proof. Let $\{W_a\}$ be a collection of subspaces of V, and let $W = \bigcap_a W_a$ be their intersection. Recall that W is defined as the set of all elements belonging to every W_a (see Appendix). Since each W_a is a subspace, each contains the zero vector. Thus the zero vector is in the intersection W, and W is non-empty. Let α and β be vectors in W and let c be a scalar. By definition of W, both α and β belong to each W_a, and because each W_a is a subspace, the vector $(c\alpha + \beta)$ is in every W_a. Thus $(c\alpha + \beta)$ is again in W. By Theorem 1, W is a subspace of V. $\blacksquare$

From Theorem 2 it follows that if S is any collection of vectors in V, then there is a smallest subspace of V which contains S, that is, a subspace which contains S and which is contained in every other subspace containing S.

Definition. *Let* S *be a set of vectors in a vector space* V. *The* **subspace spanned** *by* S *is defined to be the intersection* W *of all subspaces of* V *which contain* S. *When* S *is a finite set of vectors,* S = $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, *we shall simply call* W *the* **subspace spanned by the vectors** $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Theorem 3. *The subspace spanned by a non-empty subset* S *of a vector space* V *is the set of all linear combinations of vectors in* S.

Proof. Let W be the subspace spanned by S. Then each linear combination

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$$

of vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ in S is clearly in W. Thus W contains the set L of all linear combinations of vectors in S. The set L, on the other hand, contains S and is non-empty. If α, β belong to L then α is a linear combination,

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$$

of vectors α_i in S, and β is a linear combination,

$$\beta = y_1\beta_1 + y_2\beta_2 + \cdots + y_n\beta_n$$

of vectors β_j in S. For each scalar c,

$$c\alpha + \beta = \sum_{i=1}^{m} (cx_i)\alpha_i + \sum_{j=1}^{n} y_j\beta_j.$$

Hence $c\alpha + \beta$ belongs to L. Thus L is a subspace of V.

Now we have shown that L is a subspace of V which contains S, and also that any subspace which contains S contains L. It follows that L is the intersection of all subspaces containing S, i.e., that L is the subspace spanned by the set S. ▮

Definition. *If* $\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_k$ *are subsets of a vector space* V, *the set of all sums*

$$\alpha_1 + \alpha_2 + \cdots + \alpha_k$$

of vectors α_i *in* S_i *is called the* **sum** *of the subsets* $\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_k$ *and is denoted by*

$$\mathrm{S}_1 + \mathrm{S}_2 + \cdots + \mathrm{S}_k$$

or by

$$\sum_{i=1}^{k} \mathrm{S}_i.$$

If $W_1, W_2, \ldots, W_k$ are subspaces of V, then the sum

$$W = W_1 + W_2 + \cdots + W_k$$

is easily seen to be a subspace of V which contains each of the subspaces W_i. From this it follows, as in the proof of Theorem 3, that W is the subspace spanned by the union of $W_1, W_2, \ldots, W_k$.

EXAMPLE 8. Let F be a subfield of the field C of complex numbers. Suppose

$$\alpha_1 = (1, 2, 0, 3, 0)$$
$$\alpha_2 = (0, 0, 1, 4, 0)$$
$$\alpha_3 = (0, 0, 0, 0, 1).$$

By Theorem 3, a vector α is in the subspace W of F^5 spanned by α_1, α_2, α_3 if and only if there exist scalars c_1, c_2, c_3 in F such that

$$\alpha = c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3.$$

Thus W consists of all vectors of the form

$$\alpha = (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3)$$

where c_1, c_2, c_3 are arbitrary scalars in F. Alternatively, W can be described as the set of all 5-tuples

$$\alpha = (x_1, x_2, x_3, x_4, x_5)$$

with x_i in F such that

$$x_2 = 2x_1$$
$$x_4 = 3x_1 + 4x_3.$$

Thus $(-3, -6, 1, -5, 2)$ is in W, whereas $(2, 4, 6, 7, 8)$ is not.

EXAMPLE 9. Let F be a subfield of the field C of complex numbers, and let V be the vector space of all 2×2 matrices over F. Let W_1 be the subset of V consisting of all matrices of the form

$$\begin{bmatrix} x & y \\ z & 0 \end{bmatrix}$$

where x, y, z are arbitrary scalars in F. Finally, let W_2 be the subset of V consisting of all matrices of the form

$$\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$$

where x and y are arbitrary scalars in F. Then W_1 and W_2 are subspaces of V. Also

$$V = W_1 + W_2$$

because

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}.$$

The subspace $W_1 \cap W_2$ consists of all matrices of the form

$$\begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix}.$$

EXAMPLE 10. Let A be an $m \times n$ matrix over a field F. The **row vectors** of A are the vectors in F^n given by $\alpha_i = (A_{i1}, \ldots, A_{in})$, $i = 1, \ldots, m$. The subspace of F^n spanned by the row vectors of A is called the **row**

space of A. The subspace considered in Example 8 is the row space of the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

It is also the row space of the matrix

$$B = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -4 & -8 & 1 & -8 & 0 \end{bmatrix}.$$

EXAMPLE 11. Let V be the space of all polynomial functions over F. Let S be the subset of V consisting of the polynomial functions $f_0, f_1, f_2, \ldots$ defined by

$$f_n(x) = x^n, \qquad n = 0, 1, 2, \ldots.$$

Then V is the subspace spanned by the set S.

Exercises

1. Which of the following sets of vectors $\alpha = (a_1, \ldots, a_n)$ in R^n are subspaces of R^n $(n \geq 3)$?

(a) all α such that $a_1 \geq 0$;
(b) all α such that $a_1 + 3a_2 = a_3$;
(c) all α such that $a_2 = a_1^2$;
(d) all α such that $a_1 a_2 = 0$;
(e) all α such that a_2 is rational.

2. Let V be the (real) vector space of all functions f from R into R. Which of the following sets of functions are subspaces of V?

(a) all f such that $f(x^2) = f(x)^2$;
(b) all f such that $f(0) = f(1)$;
(c) all f such that $f(3) = 1 + f(-5)$;
(d) all f such that $f(-1) = 0$;
(e) all f which are continuous.

3. Is the vector $(3, -1, 0, -1)$ in the subspace of R^5 spanned by the vectors $(2, -1, 3, 2)$, $(-1, 1, 1, -3)$, and $(1, 1, 9, -5)$?

4. Let W be the set of all $(x_1, x_2, x_3, x_4, x_5)$ in R^5 which satisfy

$$\begin{aligned} 2x_1 - x_2 + \tfrac{4}{3}x_3 - x_4 &= 0 \\ x_1 + \tfrac{2}{3}x_3 - x_5 &= 0 \\ 9x_1 - 3x_2 + 6x_3 - 3x_4 - 3x_5 &= 0. \end{aligned}$$

Find a finite set of vectors which spans W.

5. Let F be a field and let n be a positive integer ($n \geq 2$). Let V be the vector space of all $n \times n$ matrices over F. Which of the following sets of matrices A in V are subspaces of V?

(a) all invertible A;
(b) all non-invertible A;
(c) all A such that $AB = BA$, where B is some fixed matrix in V;
(d) all A such that $A^2 = A$.

6. (a) Prove that the only subspaces of R^1 are R^1 and the zero subspace.

(b) Prove that a subspace of R^2 is R^2, or the zero subspace, or consists of all scalar multiples of some fixed vector in R^2. (The last type of subspace is, intuitively, a straight line through the origin.)

(c) Can you describe the subspaces of R^3?

7. Let W_1 and W_2 be subspaces of a vector space V such that the set-theoretic union of W_1 and W_2 is also a subspace. Prove that one of the spaces W_i is contained in the other.

8. Let V be the vector space of all functions from R into R; let V_e be the subset of even functions, $f(-x) = f(x)$; let V_o be the subset of odd functions, $f(-x) = -f(x)$.

(a) Prove that V_e and V_o are subspaces of V.
(b) Prove that $V_e + V_o = V$.
(c) Prove that $V_e \cap V_o = \{0\}$.

9. Let W_1 and W_2 be subspaces of a vector space V such that $W_1 + W_2 = V$ and $W_1 \cap W_2 = \{0\}$. Prove that for each vector α in V there are *unique* vectors α_1 in W_1 and α_2 in W_2 such that $\alpha = \alpha_1 + \alpha_2$.

2.3. Bases and Dimension

We turn now to the task of assigning a dimension to certain vector spaces. Although we usually associate 'dimension' with something geometrical, we must find a suitable algebraic definition of the dimension of a vector space. This will be done through the concept of a basis for the space.

Definition. *Let* V *be a vector space over* F. *A subset* S *of* V *is said to be* **linearly dependent** *(or simply,* **dependent***) if there exist distinct vectors* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *in* S *and scalars* $c_1, c_2, \ldots, c_n$ *in* F, *not all of which are* 0, *such that*

$$c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n = 0.$$

A set which is not linearly dependent is called **linearly independent.** *If the set* S *contains only finitely many vectors* $\alpha_1, \alpha_2, \ldots, \alpha_n$, *we sometimes say that* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *are dependent (or independent) instead of saying* S *is dependent (or independent).*

The following are easy consequences of the definition.

1. Any set which contains a linearly dependent set is linearly dependent.

2. Any subset of a linearly independent set is linearly independent.

3. Any set which contains the 0 vector is linearly dependent; for $1 \cdot 0 = 0$.

4. A set S of vectors is linearly independent if and only if each finite subset of S is linearly independent, i.e., if and only if for any distinct vectors $\alpha_1, \ldots, \alpha_n$ of S, $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$ implies each $c_i = 0$.

Definition. *Let* V *be a vector space. A* **basis** *for* V *is a linearly independent set of vectors in* V *which spans the space* V. *The space* V *is* **finite-dimensional** *if it has a finite basis.*

EXAMPLE 12. Let F be a subfield of the complex numbers. In F^3 the vectors

$$\begin{aligned} \alpha_1 &= (\ 3, 0, -3) \\ \alpha_2 &= (-1, 1, \ 2) \\ \alpha_3 &= (\ 4, 2, -2) \\ \alpha_4 &= (\ 2, 1, \ 1) \end{aligned}$$

are linearly dependent, since

$$2\alpha_1 + 2\alpha_2 - \alpha_3 + 0 \cdot \alpha_4 = 0.$$

The vectors

$$\begin{aligned} \epsilon_1 &= (1, 0, 0) \\ \epsilon_2 &= (0, 1, 0) \\ \epsilon_3 &= (0, 0, 1) \end{aligned}$$

are linearly independent

EXAMPLE 13. Let F be a field and in F^n let S be the subset consisting of the vectors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ defined by

$$\begin{aligned} \epsilon_1 &= (1, 0, 0, \ldots, 0) \\ \epsilon_2 &= (0, 1, 0, \ldots, 0) \\ &\cdots\cdots\cdots \\ \epsilon_n &= (0, 0, 0, \ldots, 1). \end{aligned}$$

Let $x_1, x_2, \ldots, x_n$ be scalars in F and put $\alpha = x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n$. Then

$$\alpha = (x_1, x_2, \ldots, x_n). \tag{2-12}$$

This shows that $\epsilon_1, \ldots, \epsilon_n$ span F^n. Since $\alpha = 0$ if and only if $x_1 = x_2 = \cdots = x_n = 0$, the vectors $\epsilon_1, \ldots, \epsilon_n$ are linearly independent. The set $S = \{\epsilon_1, \ldots, \epsilon_n\}$ is accordingly a basis for F^n. We shall call this particular basis the **standard basis** of F^n.

EXAMPLE 14. Let P be an invertible $n \times n$ matrix with entries in the field F. Then $P_1, \ldots, P_n$, the columns of P, form a basis for the space of column matrices, $F^{n\times 1}$. We see that as follows. If X is a column matrix, then

$$PX = x_1P_1 + \cdots + x_nP_n.$$

Since $PX = 0$ has only the trivial solution $X = 0$, it follows that $\{P_1, \ldots, P_n\}$ is a linearly independent set. Why does it span $F^{n\times 1}$? Let Y be any column matrix. If $X = P^{-1}Y$, then $Y = PX$, that is,

$$Y = x_1P_1 + \cdots + x_nP_n.$$

So $\{P_1, \ldots, P_n\}$ is a basis for $F^{n\times 1}$.

EXAMPLE 15. Let A be an $m \times n$ matrix and let S be the solution space for the homogeneous system $AX = 0$ (Example 7). Let R be a row-reduced echelon matrix which is row-equivalent to A. Then S is also the solution space for the system $RX = 0$. If R has r non-zero rows, then the system of equations $RX = 0$ simply expresses r of the unknowns $x_1, \ldots, x_n$ in terms of the remaining $(n - r)$ unknowns x_j. Suppose that the leading non-zero entries of the non-zero rows occur in columns $k_1, \ldots, k_r$. Let J be the set consisting of the $n - r$ indices different from $k_1, \ldots, k_r$:

$$J = \{1, \ldots, n\} - \{k_1, \ldots, k_r\}.$$

The system $RX = 0$ has the form

$$\begin{aligned} x_{k_1} + \sum_J c_{1j}x_j &= 0 \\ \vdots \qquad\quad \vdots \qquad &\quad \vdots \\ x_{k_r} + \sum_J c_{rj}x_j &= 0 \end{aligned}$$

where the c_{ij} are certain scalars. All solutions are obtained by assigning (arbitrary) values to those x_j's with j in J and computing the corresponding values of $x_{k_1}, \ldots, x_{k_r}$. For each j in J, let E_j be the solution obtained by setting $x_j = 1$ and $x_i = 0$ for all other i in J. We assert that the $(n - r)$ vectors E_j, j in J, form a basis for the solution space.

Since the column matrix E_j has a 1 in row j and zeros in the rows indexed by other elements of J, the reasoning of Example 13 shows us that the set of these vectors is linearly independent. That set spans the solution space, for this reason. If the column matrix T, with entries $t_1, \ldots, t_n$, is in the solution space, the matrix

$$N = \sum_J t_jE_j$$

is also in the solution space *and* is a solution such that $x_j = t_j$ for each j in J. The solution with that property is unique; hence, $N = T$ and T is in the span of the vectors E_j.

EXAMPLE 16. We shall now give an example of an infinite basis. Let F be a subfield of the complex numbers and let V be the space of polynomial functions over F. Recall that these functions are the functions from F into F which have a rule of the form

$$f(x) = c_0 + c_1x + \cdots + c_nx^n.$$

Let $f_k(x) = x_k$, $k = 0, 1, 2, \ldots$. The (infinite) set $\{f_0, f_1, f_2, \ldots\}$ is a basis for V. Clearly the set spans V, because the function f (above) is

$$f = c_0f_0 + c_1f_1 + \cdots + c_nf_n.$$

The reader should see that this is virtually a repetition of the definition of polynomial function, that is, a function f from F into F is a polynomial function if and only if there exists an integer n and scalars $c_0, \ldots, c_n$ such that $f = c_0f_0 + \cdots + c_nf_n$. Why are the functions independent? To show that the set $\{f_0, f_1, f_2, \ldots\}$ is independent means to show that each finite subset of it is independent. It will suffice to show that, for each n, the set $\{f_0, \ldots, f_n\}$ is independent. Suppose that

$$c_0f_0 + \cdots + c_nf_n = 0.$$

This says that

$$c_0 + c_1x + \cdots + c_nx^n = 0$$

for every x in F; in other words, every x in F is a root of the polynomial $f(x) = c_0 + c_1x + \cdots + c_nx^n$. We assume that the reader knows that a polynomial of degree n with complex coefficients cannot have more than n distinct roots. It follows that $c_0 = c_1 = \cdots = c_n = 0$.

We have exhibited an infinite basis for V. Does that mean that V is not finite-dimensional? As a matter of fact it does; however, that is not immediate from the definition, because for all we know V might also have a finite basis. That possibility is easily eliminated. (We shall eliminate it in general in the next theorem.) Suppose that we have a finite number of polynomial functions $g_1, \ldots, g_r$. There will be a largest power of x which appears (with non-zero coefficient) in $g_1(x), \ldots, g_r(x)$. If that power is k, clearly $f_{k+1}(x) = x^{k+1}$ is not in the linear span of $g_1, \ldots, g_r$. So V is not finite-dimensional.

A final remark about this example is in order. Infinite bases have nothing to do with 'infinite linear combinations.' The reader who feels an irresistible urge to inject power series

$$\sum_{k=0}^{\infty} c_kx^k$$

into this example should study the example carefully again. If that does not effect a cure, he should consider restricting his attention to finite-dimensional spaces from now on.

Theorem 4. *Let* V *be a vector space which is spanned by a finite set of vectors* $\beta_1, \beta_2, \ldots, \beta_m$. *Then any independent set of vectors in* V *is finite and contains no more than* m *elements.*

Proof. To prove the theorem it suffices to show that every subset S of V which contains more than m vectors is linearly dependent. Let S be such a set. In S there are distinct vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ where $n > m$. Since $\beta_1, \ldots, \beta_m$ span V, there exist scalars A_{ij} in F such that

$$\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i.$$

For any n scalars $x_1, x_2, \ldots, x_n$ we have

$$\begin{aligned} x_1\alpha_1 + \cdots + x_n\alpha_n &= \sum_{j=1}^{n} x_j\alpha_j \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} (A_{ij}x_j)\beta_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i. \end{aligned}$$

Since $n > m$, Theorem 6 of Chapter 1 implies that there exist scalars $x_1, x_2, \ldots, x_n$ not all 0 such that

$$\sum_{j=1}^{n} A_{ij}x_j = 0, \qquad 1 \leq i \leq m.$$

Hence $x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = 0$. This shows that S is a linearly dependent set. ∎

Corollary 1. *If* V *is a finite-dimensional vector space, then any two bases of* V *have the same (finite) number of elements.*

Proof. Since V is finite-dimensional, it has a finite basis

$$\{\beta_1, \beta_2, \ldots, \beta_m\}.$$

By Theorem 4 every basis of V is finite and contains no more than m elements. Thus if $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis, $n \leq m$. By the same argument, $m \leq n$. Hence $m = n$. ∎

This corollary allows us to define the **dimension** of a finite-dimensional vector space as the number of elements in a basis for V. We shall denote the dimension of a finite-dimensional space V by dim V. This allows us to reformulate Theorem 4 as follows.

Corollary 2. *Let* V *be a finite-dimensional vector space and let* n = *dim* V. *Then*

(a) *any subset of* V *which contains more than* n *vectors is linearly dependent;*

(b) *no subset of* V *which contains fewer than* n *vectors can span* V.

EXAMPLE 17. If F is a field, the dimension of F^n is n, because the standard basis for F^n contains n vectors. The matrix space $F^{m\times n}$ has dimension mn. That should be clear by analogy with the case of F^n, because the mn matrices which have a 1 in the i, j place with zeros elsewhere form a basis for $F^{m\times n}$. If A is an $m \times n$ matrix, then the solution space for A has dimension $n - r$, where r is the number of non-zero rows in a row-reduced echelon matrix which is row-equivalent to A. See Example 15.

If V is any vector space over F, the zero subspace of V is spanned by the vector 0, but $\{0\}$ is a linearly dependent set and not a basis. For this reason, we shall agree that the zero subspace has dimension 0. Alternatively, we could reach the same conclusion by arguing that the empty set is a basis for the zero subspace. The empty set spans $\{0\}$, because the intersection of all subspaces containing the empty set is $\{0\}$, and the empty set is linearly independent because it contains no vectors.

Lemma. *Let* S *be a linearly independent subset of a vector space* V. *Suppose* β *is a vector in* V *which is not in the subspace spanned by* S. *Then the set obtained by adjoining* β *to* S *is linearly independent.*

Proof. Suppose $\alpha_1, \ldots, \alpha_m$ are distinct vectors in S and that

$$c_1\alpha_1 + \cdots + c_m\alpha_m + b\beta = 0.$$

Then $b = 0$; for otherwise,

$$\beta = \left(-\frac{c_1}{b}\right)\alpha_1 + \cdots + \left(-\frac{c_m}{b}\right)\alpha_m$$

and β is in the subspace spanned by S. Thus $c_1\alpha_1 + \cdots + c_m\alpha_m = 0$, and since S is a linearly independent set each $c_i = 0$. ∎

Theorem 5. *If* W *is a subspace of a finite-dimensional vector space* V, *every linearly independent subset of* W *is finite and is part of a (finite) basis for* W.

Proof. Suppose S_0 is a linearly independent subset of W. If S is a linearly independent subset of W containing S_0, then S is also a linearly independent subset of V; since V is finite-dimensional, S contains no more than dim V elements.

We extend S_0 to a basis for W, as follows. If S_0 spans W, then S_0 is a basis for W and we are done. If S_0 does not span W, we use the preceding lemma to find a vector β_1 in W such that the set $S_1 = S_0 \cup \{\beta_1\}$ is independent. If S_1 spans W, fine. If not, apply the lemma to obtain a vector β_2

in W such that $S_2 = S_1 \cup \{\beta_2\}$ is independent. If we continue in this way, then (in not more than dim V steps) we reach a set

$$S_m = S_0 \cup \{\beta_1, \ldots, \beta_m\}$$

which is a basis for W. ∎

Corollary 1. *If* W *is a proper subspace of a finite-dimensional vector space* V, *then* W *is finite-dimensional and dim* W $<$ *dim* V.

Proof. We may suppose W contains a vector $\alpha \neq 0$. By Theorem 5 and its proof, there is a basis of W containing α which contains no more than dim V elements. Hence W is finite-dimensional, and dim $W \leq$ dim V. Since W is a proper subspace, there is a vector β in V which is not in W. Adjoining β to any basis of W, we obtain a linearly independent subset of V. Thus dim $W <$ dim V. ∎

Corollary 2. *In a finite-dimensional vector space* V *every non-empty linearly independent set of vectors is part of a basis.*

Corollary 3. *Let* A *be an* n × n *matrix over a field* F, *and suppose the row vectors of* A *form a linearly independent set of vectors in* F^n. *Then* A *is invertible.*

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the row vectors of A, and suppose W is the subspace of F^n spanned by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Since $\alpha_1, \alpha_2, \ldots, \alpha_n$ are linearly independent, the dimension of W is n. Corollary 1 now shows that $W = F^n$. Hence there exist scalars B_{ij} in F such that

$$\epsilon_i = \sum_{j=1}^{n} B_{ij}\alpha_j, \qquad 1 \leq i \leq n$$

where $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is the standard basis of F^n. Thus for the matrix B with entries B_{ij} we have

$$BA = I. \quad ∎$$

Theorem 6. *If* W_1 *and* W_2 *are finite-dimensional subspaces of a vector space* V, *then* $W_1 + W_2$ *is finite-dimensional and*

$$\textit{dim}\ W_1 + \textit{dim}\ W_2 = \textit{dim}\ (W_1 \cap W_2) + \textit{dim}\ (W_1 + W_2).$$

Proof. By Theorem 5 and its corollaries, $W_1 \cap W_2$ has a finite basis $\{\alpha_1, \ldots, \alpha_k\}$ which is part of a basis

$$\{\alpha_1, \ldots, \alpha_k, \quad \beta_1, \ldots, \beta_m\} \quad \text{for} \quad W_1$$

and part of a basis

$$\{\alpha_1, \ldots, \alpha_k, \quad \gamma_1, \ldots, \gamma_n\} \quad \text{for} \quad W_2.$$

The subspace $W_1 + W_2$ is spanned by the vectors

$$\alpha_1, \ldots, \alpha_k, \qquad \beta_1, \ldots, \beta_m, \qquad \gamma_1, \ldots, \gamma_n$$

and these vectors form an independent set. For suppose

$$\Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j + \Sigma\, z_r\gamma_r = 0.$$

Then

$$-\Sigma\, z_r\gamma_r = \Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j$$

which shows that $\Sigma\, z_r\gamma_r$ belongs to W_1. As $\Sigma\, z_r\gamma_r$ also belongs to W_2 it follows that

$$\Sigma\, z_r\gamma_r = \Sigma\, c_i\alpha_i$$

for certain scalars $c_1, \ldots, c_k$. Because the set

$$\{\alpha_1, \ldots, \alpha_k, \quad \gamma_1, \ldots, \gamma_n\}$$

is independent, each of the scalars $z_r = 0$. Thus

$$\Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j = 0$$

and since

$$\{\alpha_1, \ldots, \alpha_k, \quad \beta_1, \ldots, \beta_m\}$$

is also an independent set, each $x_i = 0$ and each $y_j = 0$. Thus,

$$\{\alpha_1, \ldots, \alpha_k, \quad \beta_1, \ldots, \beta_m, \quad \gamma_1, \ldots, \gamma_n\}$$

is a basis for $W_1 + W_2$. Finally

$$\begin{aligned} \dim W_1 + \dim W_2 &= (k + m) + (k + n) \\ &= k + (m + k + n) \\ &= \dim (W_1 \cap W_2) + \dim (W_1 + W_2). \end{aligned}$$ ∎

Let us close this section with a remark about linear independence and dependence. We defined these concepts for sets of vectors. It is useful to have them defined for finite sequences (ordered n-tuples) of vectors: $\alpha_1, \ldots, \alpha_n$. We say that the vectors $\alpha_1, \ldots, \alpha_n$ are **linearly dependent** if there exist scalars $c_1, \ldots, c_n$, not all 0, such that $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$. This is all so natural that the reader may find that he has been using this terminology already. What is the difference between a finite sequence $\alpha_1, \ldots, \alpha_n$ and a set $\{\alpha_1, \ldots, \alpha_n\}$? There are two differences, identity and order.

If we discuss the set $\{\alpha_1, \ldots, \alpha_n\}$, usually it is presumed that no two of the vectors $\alpha_1, \ldots, \alpha_n$ are identical. In a sequence $\alpha_1, \ldots, \alpha_n$ all the α_i's may be the same vector. If $\alpha_i = \alpha_j$ for some $i \neq j$, then the sequence $\alpha_1, \ldots, \alpha_n$ is linearly dependent:

$$\alpha_i + (-1)\alpha_j = 0.$$

Thus, if $\alpha_1, \ldots, \alpha_n$ are linearly independent, they are distinct and we may talk about the set $\{\alpha_1, \ldots, \alpha_n\}$ and know that it has n vectors in it. So, clearly, no confusion will arise in discussing bases and dimension. The dimension of a finite-dimensional space V is the largest n such that some n-tuple of vectors in V is linearly independent—and so on. The reader

who feels that this paragraph is much ado about nothing might ask himself whether the vectors

$$\alpha_1 = (e^{\pi/2}, 1)$$
$$\alpha_2 = (\sqrt[3]{110}, 1)$$

are linearly independent in R^2.

The elements of a sequence are enumerated in a specific order. A set is a collection of objects, with no specified arrangement or order. Of course, to describe the set we may list its members, and that requires choosing an order. But, the order is not part of the set. The sets $\{1, 2, 3, 4\}$ and $\{4, 3, 2, 1\}$ are identical, whereas 1, 2, 3, 4 is quite a different sequence from 4, 3, 2, 1. The order aspect of sequences has no bearing on questions of independence, dependence, etc., because dependence (as defined) is not affected by the order. The sequence $\alpha_n, \ldots, \alpha_1$ is dependent if and only if the sequence $\alpha_1, \ldots, \alpha_n$ is dependent. In the next section, order will be important.

Exercises

1. Prove that if two vectors are linearly dependent, one of them is a scalar multiple of the other.

2. Are the vectors

$$\alpha_1 = (1, 1, 2, 4), \quad \alpha_2 = (2, -1, -5, 2)$$
$$\alpha_3 = (1, -1, -4, 0), \quad \alpha_4 = (2, 1, 1, 6)$$

linearly independent in R^4?

3. Find a basis for the subspace of R^4 spanned by the four vectors of Exercise 2.

4. Show that the vectors

$$\alpha_1 = (1, 0, -1), \quad \alpha_2 = (1, 2, 1), \quad \alpha_3 = (0, -3, 2)$$

form a basis for R^3. Express each of the standard basis vectors as linear combinations of α_1, α_2, and α_3.

5. Find three vectors in R^3 which are linearly dependent, and are such that any two of them are linearly independent.

6. Let V be the vector space of all 2×2 matrices over the field F. Prove that V has dimension 4 by exhibiting a basis for V which has four elements.

7. Let V be the vector space of Exercise 6. Let W_1 be the set of matrices of the form

$$\begin{bmatrix} x & -x \\ y & z \end{bmatrix}$$

and let W_2 be the set of matrices of the form

$$\begin{bmatrix} a & b \\ -a & c \end{bmatrix}.$$

(a) Prove that W_1 and W_2 are subspaces of V.
(b) Find the dimensions of W_1, W_2, $W_1 + W_2$, and $W_1 \cap W_2$.

8. Again let V be the space of 2×2 matrices over F. Find a basis $\{A_1, A_2, A_3, A_4\}$ for V such that $A_j^2 = A_j$ for each j.

9. Let V be a vector space over a subfield F of the complex numbers. Suppose α, β, and γ are linearly independent vectors in V. Prove that $(\alpha + \beta)$, $(\beta + \gamma)$, and $(\gamma + \alpha)$ are linearly independent.

10. Let V be a vector space over the field F. Suppose there are a finite number of vectors $\alpha_1, \ldots, \alpha_r$ in V which span V. Prove that V is finite-dimensional.

11. Let V be the set of all 2×2 matrices A with *complex* entries which satisfy $A_{11} + A_{22} = 0$.

(a) Show that V is a vector space over the field of *real* numbers, with the usual operations of matrix addition and multiplication of a matrix by a scalar.
(b) Find a basis for this vector space.
(c) Let W be the set of all matrices A in V such that $A_{21} = -\overline{A}_{12}$ (the bar denotes complex conjugation). Prove that W is a subspace of V and find a basis for W.

12. Prove that the space of all $m \times n$ matrices over the field F has dimension mn, by exhibiting a basis for this space.

13. Discuss Exercise 9, when V is a vector space over the field with two elements described in Exercise 5, Section 1.1.

14. Let V be the set of real numbers. Regard V as a vector space over the field of *rational* numbers, with the usual operations. Prove that this vector space is *not* finite-dimensional.

2.4. Coordinates

One of the useful features of a basis $\mathcal{B}$ in an n-dimensional space V is that it essentially enables one to introduce coordinates in V analogous to the 'natural coordinates' x_i of a vector $\alpha = (x_1, \ldots, x_n)$ in the space F^n. In this scheme, the coordinates of a vector α in V relative to the basis $\mathcal{B}$ will be the scalars which serve to express α as a linear combination of the vectors in the basis. Thus, we should like to regard the natural coordinates of a vector α in F^n as being defined by α and the standard basis for F^n; however, in adopting this point of view we must exercise a certain amount of care. If

$$\alpha = (x_1, \ldots, x_n) = \Sigma\, x_i \epsilon_i$$

and $\mathcal{B}$ is the standard basis for F^n, just how are the coordinates of α determined by $\mathcal{B}$ and α? One way to phrase the answer is this. A given vector α has a unique expression as a linear combination of the standard basis vectors, and the ith coordinate x_i of α is the coefficient of ϵ_i in this expression. From this point of view we are able to say which is the ith coordinate

because we have a 'natural' ordering of the vectors in the standard basis, that is, we have a rule for determining which is the 'first' vector in the basis, which is the 'second,' and so on. If $\mathfrak{B}$ is an arbitrary basis of the n-dimensional space V, we shall probably have no natural ordering of the vectors in $\mathfrak{B}$, and it will therefore be necessary for us to impose some order on these vectors before we can define 'the ith coordinate of α relative to $\mathfrak{B}$.' To put it another way, coordinates will be defined relative to sequences of vectors rather than sets of vectors.

Definition. *If* V *is a finite-dimensional vector space, an* **ordered basis** *for* V *is a finite sequence of vectors which is linearly independent and spans* V.

If the sequence $\alpha_1, \ldots, \alpha_n$ is an ordered basis for V, then the set $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V. The ordered basis is the set, together with the specified ordering. We shall engage in a slight abuse of notation and describe all that by saying that

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$$

is an ordered basis for V.

Now suppose V is a finite-dimensional vector space over the field F and that

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$$

is an ordered basis for V. Given α in V, there is a unique n-tuple $(x_1, \ldots, x_n)$ of scalars such that

$$\alpha = \sum_{i=1}^{n} x_i\alpha_i.$$

The n-tuple is unique, because if we also have

$$\alpha = \sum_{i=1}^{n} z_i\alpha_i$$

then

$$\sum_{i=1}^{n} (x_i - z_i)\alpha_i = 0$$

and the linear independence of the α_i tells us that $x_i - z_i = 0$ for each i. We shall call x_i the ith **coordinate of α relative to the ordered basis**

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}.$$

If

$$\beta = \sum_{i=1}^{n} y_i\alpha_i$$

then

$$\alpha + \beta = \sum_{i=1}^{n} (x_i + y_i)\alpha_i$$

so that the ith coordinate of $(\alpha + \beta)$ in this ordered basis is $(x_i + y_i)$.

Similarly, the ith coordinate of $(c\alpha)$ is cx_i. One should also note that every n-tuple $(x_1, \ldots, x_n)$ in F^n is the n-tuple of coordinates of some vector in V, namely the vector

$$\sum_{i=1}^{n} x_i\alpha_i.$$

To summarize, each ordered basis for V determines a one-one correspondence

$$\alpha \to (x_1, \ldots, x_n)$$

between the set of all vectors in V and the set of all n-tuples in F^n. This correspondence has the property that the correspondent of $(\alpha + \beta)$ is the sum in F^n of the correspondents of α and β, and that the correspondent of $(c\alpha)$ is the product in F^n of the scalar c and the correspondent of α.

One might wonder at this point why we do not simply select some ordered basis for V and describe each vector in V by its corresponding n-tuple of coordinates, since we would then have the convenience of operating only with n-tuples. This would defeat our purpose, for two reasons. First, as our axiomatic definition of vector space indicates, we are attempting to learn to reason with vector spaces as abstract algebraic systems. Second, even in those situations in which we use coordinates, the significant results follow from our ability to change the coordinate system, i.e., to change the ordered basis.

Frequently, it will be more convenient for us to use the **coordinate matrix of α relative to the ordered basis $\mathcal{B}$**:

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

rather than the n-tuple $(x_1, \ldots, x_n)$ of coordinates. To indicate the dependence of this coordinate matrix on the basis, we shall use the symbol

$$[\alpha]_{\mathcal{B}}$$

for the coordinate matrix of the vector α relative to the ordered basis $\mathcal{B}$. This notation will be particularly useful as we now proceed to describe what happens to the coordinates of a vector α as we change from one ordered basis to another.

Suppose then that V is n-dimensional and that

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{and} \quad \mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

are two ordered bases for V. There are unique scalars P_{ij} such that

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i, \qquad 1 \leq j \leq n. \tag{2-13}$$

Let $x'_1, \ldots, x'_n$ be the coordinates of a given vector α in the ordered basis $\mathcal{B}'$. Then

$$\begin{aligned}\alpha &= x_1'\alpha_1' + \cdots + x_n'\alpha_n' \\ &= \sum_{j=1}^{n} x_j'\alpha_j' \\ &= \sum_{j=1}^{n} x_j' \sum_{i=1}^{n} P_{ij}\alpha_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} (P_{ij}x_j')\alpha_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x_j' \right) \alpha_i.\end{aligned}$$

Thus we obtain the relation

$$\alpha = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x_j' \right) \alpha_i. \tag{2-14}$$

Since the coordinates $x_1, x_2, \ldots, x_n$ of α in the ordered basis $\mathfrak{B}$ are uniquely determined, it follows from (2-14) that

$$x_i = \sum_{j=1}^{n} P_{ij}x_j', \qquad 1 \leq i \leq n. \tag{2-15}$$

Let P be the $n \times n$ matrix whose i, j entry is the scalar P_{ij}, and let X and X' be the coordinate matrices of the vector α in the ordered bases $\mathfrak{B}$ and $\mathfrak{B}'$. Then we may reformulate (2-15) as

$$X = PX'. \tag{2-16}$$

Since $\mathfrak{B}$ and $\mathfrak{B}'$ are linearly independent sets, $X = 0$ if and only if $X' = 0$. Thus from (2-16) and Theorem 7 of Chapter 1, it follows that P is invertible. Hence

$$X' = P^{-1}X. \tag{2-17}$$

If we use the notation introduced above for the coordinate matrix of a vector relative to an ordered basis, then (2-16) and (2-17) say

$$\begin{aligned}[\alpha]_{\mathfrak{B}} &= P[\alpha]_{\mathfrak{B}'} \\ [\alpha]_{\mathfrak{B}'} &= P^{-1}[\alpha]_{\mathfrak{B}}.\end{aligned}$$

Thus the preceding discussion may be summarized as follows.

Theorem 7. *Let* V *be an* n*-dimensional vector space over the field* F, *and let* $\mathfrak{B}$ *and* $\mathfrak{B}'$ *be two ordered bases of* V. *Then there is a unique, necessarily invertible,* n × n *matrix* P *with entries in* F *such that*

$$\text{(i)} \qquad [\alpha]_{\mathfrak{B}} = \mathrm{P}[\alpha]_{\mathfrak{B}'}$$
$$\text{(ii)} \qquad [\alpha]_{\mathfrak{B}'} = \mathrm{P}^{-1}[\alpha]_{\mathfrak{B}}$$

for every vector α *in* V. *The columns of* P *are given by*

$$\mathrm{P}_j = [\alpha_j']_{\mathfrak{B}}, \qquad \mathrm{j} = 1, \ldots, \mathrm{n}.$$

To complete the above analysis we shall also prove the following result.

Theorem 8. *Suppose* P *is an* n × n *invertible matrix over* F. *Let* V *be an* n*-dimensional vector space over* F, *and let* $\mathcal{B}$ *be an ordered basis of* V. *Then there is a unique ordered basis* $\mathcal{B}'$ *of* V *such that*

(i) $$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$

(ii) $$[\alpha]_{\mathcal{B}'} = P^{-1}[\alpha]_{\mathcal{B}}$$

for every vector α *in* V.

Proof. Let $\mathcal{B}$ consist of the vectors $\alpha_1, \ldots, \alpha_n$. If $\mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$ is an ordered basis of V for which (i) is valid, it is clear that

$$\alpha_j' = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Thus we need only show that the vectors α_j', defined by these equations, form a basis. Let $Q = P^{-1}$. Then

$$\begin{aligned}\sum_j Q_{jk}\alpha_j' &= \sum_j Q_{jk} \sum_i P_{ij}\alpha_i \\ &= \sum_j \sum_i P_{ij}Q_{jk}\,\alpha_i \\ &= \sum_i \left(\sum_j P_{ij}Q_{jk}\right)\alpha_i \\ &= \alpha_k.\end{aligned}$$

Thus the subspace spanned by the set

$$\mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$$

contains $\mathcal{B}$ and hence equals V. Thus $\mathcal{B}'$ is a basis, and from its definition and Theorem 7, it is clear that (i) is valid and hence also (ii). ∎

EXAMPLE 18. Let F be a field and let

$$\alpha = (x_1, x_2, \ldots, x_n)$$

be a vector in F^n. If $\mathcal{B}$ is the standard ordered basis of F^n,

$$\mathcal{B} = \{\epsilon_1, \ldots, \epsilon_n\}$$

the coordinate matrix of the vector α in the basis $\mathcal{B}$ is given by

$$[\alpha]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

EXAMPLE 19. Let R be the field of the real numbers and let θ be a fixed real number. The matrix

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is invertible with inverse,

$$P^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Thus for each θ the set $\mathcal{B}'$ consisting of the vectors $(\cos\theta, \sin\theta)$, $(-\sin\theta, \cos\theta)$ is a basis for R^2; intuitively this basis may be described as the one obtained by rotating the standard basis through the angle θ. If α is the vector (x_1, x_2), then

$$[\alpha]_{\mathcal{B}'} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

or

$$\begin{aligned} x_1' &= x_1 \cos\theta + x_2 \sin\theta \\ x_2' &= -x_1 \sin\theta + x_2 \cos\theta. \end{aligned}$$

EXAMPLE 20. Let F be a subfield of the complex numbers. The matrix

$$P = \begin{bmatrix} -1 & 4 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix}$$

is invertible with inverse

$$P^{-1} = \begin{bmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{bmatrix}.$$

Thus the vectors

$$\begin{aligned} \alpha_1' &= (-1, 0, 0) \\ \alpha_2' &= (4, 2, 0) \\ \alpha_3' &= (5, -3, 8) \end{aligned}$$

form a basis $\mathcal{B}'$ of F^3. The coordinates x_1', x_2', x_3' of the vector $\alpha = (x_1, x_2, x_3)$ in the basis $\mathcal{B}'$ are given by

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} -x_1 + 2x_2 + \frac{11}{8}x_3 \\ \frac{1}{2}x_2 + \frac{3}{16}x_3 \\ \frac{1}{8}x_3 \end{bmatrix} = \begin{bmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

In particular,

$$(3, 2, -8) = -10\alpha_1' - \tfrac{1}{2}\alpha_2' - \alpha_3'.$$

Exercises

1. Show that the vectors

$$\begin{aligned} \alpha_1 &= (1, 1, 0, 0), & \alpha_2 &= (0, 0, 1, 1) \\ \alpha_3 &= (1, 0, 0, 4), & \alpha_4 &= (0, 0, 0, 2) \end{aligned}$$

form a basis for R^4. Find the coordinates of each of the standard basis vectors in the ordered basis $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$.

2. Find the coordinate matrix of the vector $(1, 0, 1)$ in the basis of C^3 consisting of the vectors $(2i, 1, 0)$, $(2, -1, 1)$, $(0, 1 + i, 1 - i)$, in that order.

3. Let $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ be the ordered basis for R^3 consisting of

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 1, 1), \qquad \alpha_3 = (1, 0, 0).$$

What are the coordinates of the vector (a, b, c) in the ordered basis $\mathcal{B}$?

4. Let W be the subspace of C^3 spanned by $\alpha_1 = (1, 0, i)$ and $\alpha_2 = (1 + i, 1, -1)$.

(a) Show that α_1 and α_2 form a basis for W.

(b) Show that the vectors $\beta_1 = (1, 1, 0)$ and $\beta_2 = (1, i, 1 + i)$ are in W and form another basis for W.

(c) What are the coordinates of α_1 and α_2 in the ordered basis $\{\beta_1, \beta_2\}$ for W?

5. Let $\alpha = (x_1, x_2)$ and $\beta = (y_1, y_2)$ be vectors in R^2 such that

$$x_1y_1 + x_2y_2 = 0, \qquad x_1^2 + x_2^2 = y_1^2 + y_2^2 = 1.$$

Prove that $\mathcal{B} = \{\alpha, \beta\}$ is a basis for R^2. Find the coordinates of the vector (a, b) in the ordered basis $\mathcal{B} = \{\alpha, \beta\}$. (The conditions on α and β say, geometrically, that α and β are perpendicular and each has length 1.)

6. Let V be the vector space over the complex numbers of all functions from R into C, i.e., the space of all complex-valued functions on the real line. Let $f_1(x) = 1$, $f_2(x) = e^{ix}$, $f_3(x) = e^{-ix}$.

(a) Prove that f_1, f_2, and f_3 are linearly independent.

(b) Let $g_1(x) = 1$, $g_2(x) = \cos x$, $g_3(x) = \sin x$. Find an invertible 3×3 matrix P such that

$$g_j = \sum_{i=1}^{3} P_{ij} f_i.$$

7. Let V be the (real) vector space of all polynomial functions from R into R of degree 2 or less, i.e., the space of all functions f of the form

$$f(x) = c_0 + c_1x + c_2x^2.$$

Let t be a fixed real number and define

$$g_1(x) = 1, \qquad g_2(x) = x + t, \qquad g_3(x) = (x + t)^2.$$

Prove that $\mathcal{B} = \{g_1, g_2, g_3\}$ is a basis for V. If

$$f(x) = c_0 + c_1x + c_2x^2$$

what are the coordinates of f in this ordered basis $\mathcal{B}$?

2.5. *Summary of Row-Equivalence*

In this section we shall utilize some elementary facts on bases and dimension in finite-dimensional vector spaces to complete our discussion of row-equivalence of matrices. We recall that if A is an $m \times n$ matrix over the field F the row vectors of A are the vectors $\alpha_1, \ldots, \alpha_m$ in F^n defined by

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

and that the row space of A is the subspace of F^n spanned by these vectors. The **row rank** of A is the dimension of the row space of A.

If P is a $k \times m$ matrix over F, then the product $B = PA$ is a $k \times n$ matrix whose row vectors $\beta_1, \ldots, \beta_k$ are linear combinations

$$\beta_i = P_{i1}\alpha_1 + \cdots + P_{im}\alpha_m$$

of the row vectors of A. Thus the row space of B is a subspace of the row space of A. If P is an $m \times m$ invertible matrix, then B is row-equivalent to A so that the symmetry of row-equivalence, or the equation $A = P^{-1}B$, implies that the row space of A is also a subspace of the row space of B.

Theorem 9. *Row-equivalent matrices have the same row space.*

Thus we see that to study the row space of A we may as well study the row space of a row-reduced echelon matrix which is row-equivalent to A. This we proceed to do.

Theorem 10. *Let* R *be a non-zero row-reduced echelon matrix. Then the non-zero row vectors of* R *form a basis for the row space of* R.

Proof. Let $\rho_1, \ldots, \rho_r$ be the non-zero row vectors of R:

$$\rho_i = (R_{i1}, \ldots, R_{in}).$$

Certainly these vectors span the row space of R; we need only prove they are linearly independent. Since R is a row-reduced echelon matrix, there are positive integers $k_1, \ldots, k_r$ such that, for $i \leq r$

$$\text{(2-18)} \qquad \begin{array}{ll} \text{(a)} & R(i, j) = 0 \quad \text{if} \quad j < k_i \\ \text{(b)} & R(i, k_j) = \delta_{ij} \\ \text{(c)} & k_1 < \cdots < k_r. \end{array}$$

Suppose $\beta = (b_1, \ldots, b_n)$ is a vector in the row space of R:

$$\text{(2-19)} \qquad \beta = c_1\rho_1 + \cdots + c_r\rho_r.$$

Then we claim that $c_j = b_{k_j}$. For, by (2-18)

$$\text{(2-20)} \qquad \begin{aligned} b_{k_j} &= \sum_{i=1}^{r} c_i R(i, k_j) \\ &= \sum_{i=1}^{r} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

In particular, if $\beta = 0$, i.e., if $c_1\rho_1 + \cdots + c_r\rho_r = 0$, then c_j must be the k_jth coordinate of the zero vector so that $c_j = 0$, $j = 1, \ldots, r$. Thus $\rho_1, \ldots, \rho_r$ are linearly independent. ∎

Theorem 11. *Let* m *and* n *be positive integers and let* F *be a field. Suppose* W *is a subspace of* F^n *and dim* W $\leq$ m. *Then there is precisely one* m $\times$ n *row-reduced echelon matrix over* F *which has* W *as its row space.*

Proof. There is at least one $m \times n$ row-reduced echelon matrix with row space W. Since dim $W \leq m$, we can select some m vectors $\alpha_1, \ldots, \alpha_m$ in W which span W. Let A be the $m \times n$ matrix with row vectors $\alpha_1, \ldots, \alpha_m$ and let R be a row-reduced echelon matrix which is row-equivalent to A. Then the row space of R is W.

Now let R be any row-reduced echelon matrix which has W as its row space. Let $\rho_1, \ldots, \rho_r$ be the non-zero row vectors of R and suppose that the leading non-zero entry of ρ_i occurs in column k_i, $i = 1, \ldots, r$. The vectors $\rho_1, \ldots, \rho_r$ form a basis for W. In the proof of Theorem 10, we observed that if $\beta = (b_1, \ldots, b_n)$ is in W, then

$$\beta = c_1\rho_1 + \cdots + c_r\rho_r,$$

and $c_i = b_{k_i}$; in other words, the unique expression for β as a linear combination of $\rho_1, \ldots, \rho_r$ is

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i. \tag{2-21}$$

Thus any vector β is determined if one knows the coordinates b_{k_i}, $i = 1, \ldots, r$. For example, ρ_s is the unique vector in W which has k_sth coordinate 1 and k_ith coordinate 0 for $i \neq s$.

Suppose β is in W and $\beta \neq 0$. We claim the first non-zero coordinate of β occurs in one of the columns k_s. Since

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i$$

and $\beta \neq 0$, we can write

$$\beta = \sum_{i=s}^{r} b_{k_i}\rho_i, \qquad b_{k_s} \neq 0. \tag{2-22}$$

From the conditions (2-18) one has $R_{ij} = 0$ if $i > s$ and $j \leq k_s$. Thus

$$\beta = (0, \ldots, 0, \quad b_{k_s}, \ldots, b_n), \qquad b_{k_s} \neq 0$$

and the first non-zero coordinate of β occurs in column k_s. Note also that for each k_s, $s = 1, \ldots, r$, there exists a vector in W which has a non-zero k_sth coordinate, namely ρ_s.

It is now clear that R is uniquely determined by W. The description of R in terms of W is as follows. We consider all vectors $\beta = (b_1, \ldots, b_n)$ in W. If $\beta \neq 0$, then the first non-zero coordinate of β must occur in some column t:

$$\beta = (0, \ldots, 0, \quad b_t, \ldots, b_n), \qquad b_t \neq 0.$$

Let $k_1, \ldots, k_r$ be those positive integers t such that there is some $\beta \neq 0$ in W, the first non-zero coordinate of which occurs in column t. Arrange $k_1, \ldots, k_r$ in the order $k_1 < k_2 < \cdots < k_r$. For each of the positive integers k_s there will be one and only one vector ρ_s in W such that the k_sth coordinate of ρ_s is 1 and the k_ith coordinate of ρ_s is 0 for $i \neq s$. Then R is the $m \times n$ matrix which has row vectors $\rho_1, \ldots, \rho_r, 0, \ldots, 0$. ∎

Corollary. *Each* m × n *matrix* A *is row-equivalent to one and only one row-reduced echelon matrix.*

Proof. We know that A is row-equivalent to at least one row-reduced echelon matrix R. If A is row-equivalent to another such matrix R', then R is row-equivalent to R'; hence, R and R' have the same row space and must be identical. ▮

Corollary. *Let* A *and* B *be* m × n *matrices over the field* F. *Then* A *and* B *are row-equivalent if and only if they have the same row space.*

Proof. We know that if A and B are row-equivalent, then they have the same row space. So suppose that A and B have the same row space. Now A is row-equivalent to a row-reduced echelon matrix R and B is row-equivalent to a row-reduced echelon matrix R'. Since A and B have the same row space, R and R' have the same row space. Thus $R = R'$ and A is row-equivalent to B. ▮

To summarize—if A and B are $m \times n$ matrices over the field F, the following statements are equivalent:

1. A and B are row-equivalent.
2. A and B have the same row space.
3. $B = PA$, where P is an invertible $m \times m$ matrix.

A fourth equivalent statement is that the homogeneous systems $AX = 0$ and $BX = 0$ have the same solutions; however, although we know that the row-equivalence of A and B implies that these systems have the same solutions, it seems best to leave the proof of the converse until later.

2.6. Computations Concerning Subspaces

We should like now to show how elementary row operations provide a standardized method of answering certain concrete questions concerning subspaces of F^n. We have already derived the facts we shall need. They are gathered here for the convenience of the reader. The discussion applies to any n-dimensional vector space over the field F, if one selects a fixed ordered basis $\mathcal{B}$ and describes each vector α in V by the n-tuple $(x_1, \ldots, x_n)$ which gives the coordinates of α in the ordered basis $\mathcal{B}$.

Suppose we are given m vectors $\alpha_1, \ldots, \alpha_m$ in F^n. We consider the following questions.

1. How does one determine if the vectors $\alpha_1, \ldots, \alpha_m$ are linearly independent? More generally, how does one find the dimension of the subspace W spanned by these vectors?

2. Given β in F^n, how does one determine whether β is a linear combination of $\alpha_1, \ldots, \alpha_m$, i.e., whether β is in the subspace W?

3. How can one give an explicit description of the subspace W?

The third question is a little vague, since it does not specify what is meant by an 'explicit description'; however, we shall clear up this point by giving the sort of description we have in mind. With this description, questions (1) and (2) can be answered immediately.

Let A be the $m \times n$ matrix with row vectors α_i:

$$\alpha_i = (A_{i1}, \ldots, A_{in}).$$

Perform a sequence of elementary row operations, starting with A and terminating with a row-reduced echelon matrix R. We have previously described how to do this. At this point, the dimension of W (the row space of A) is apparent, since this dimension is simply the number of non-zero row vectors of R. If $\rho_1, \ldots, \rho_r$ are the non-zero row vectors of R, then $\mathfrak{B} = \{\rho_1, \ldots, \rho_r\}$ is a basis for W. If the first non-zero coordinate of ρ_i is the k_ith one, then we have for $i \leq r$

(a) $$R(i, j) = 0, \quad \text{if} \quad j < k_i$$

(b) $$R(i, k_j) = \delta_{ij}$$

(c) $$k_1 < \cdots < k_r.$$

The subspace W consists of all vectors

$$\beta = c_1\rho_1 + \cdots + c_r\rho_r$$

$$= \sum_{i=1}^{r} c_i(R_{i1}, \ldots, R_{in}).$$

The coordinates $b_1, \ldots, b_n$ of such a vector β are then

$$b_j = \sum_{i=1}^{r} c_i R_{ij}. \tag{2-23}$$

In particular, $b_{k_i} = c_j$, and so if $\beta = (b_1, \ldots, b_n)$ is a linear combination of the ρ_i, it must be the particular linear combination

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i. \tag{2-24}$$

The conditions on β that (2-24) should hold are

$$b_j = \sum_{i=1}^{r} b_{k_i} R_{ij}, \qquad j = 1, \ldots, n. \tag{2-25}$$

Now (2-25) is the explicit description of the subspace W spanned by $\alpha_1, \ldots, \alpha_m$, that is, the subspace consists of all vectors β in F^n whose coordinates satisfy (2-25). What kind of description is (2-25)? In the first place it describes W as all solutions $\beta = (b_1, \ldots, b_n)$ of the system of homogeneous linear equations (2-25). This system of equations is of a very special nature, because it expresses $(n - r)$ of the coordinates as

linear combinations of the r distinguished coordinates $b_{k_1}, \ldots, b_{k_r}$. One has complete freedom of choice in the coordinates b_{k_i}, that is, if $c_1, \ldots, c_r$ are *any* r scalars, there is one and only one vector β in W which has c_i as its k_ith coordinate.

The significant point here is this: Given the vectors α_i, row-reduction is a straightforward method of determining the integers $r, k_1, \ldots, k_r$ and the scalars R_{ij} which give the description (2-25) of the subspace spanned by $\alpha_1, \ldots, \alpha_m$. One should observe as we did in Theorem 11 that every subspace W of F^n has a description of the type (2-25). We should also point out some things about question (2). We have already stated how one can find an invertible $m \times m$ matrix P such that $R = PA$, in Section 1.4. The knowledge of P enables one to find the scalars $x_1, \ldots, x_m$ such that

$$\beta = x_1\alpha_1 + \cdots + x_m\alpha_m$$

when this is possible. For the row vectors of R are given by

$$\rho_i = \sum_{j=1}^{m} P_{ij}\alpha_j$$

so that if β is a linear combination of the α_j, we have

$$\begin{aligned}\beta &= \sum_{i=1}^{r} b_{k_i}\rho_i \\ &= \sum_{i=1}^{r} b_{k_i} \sum_{j=1}^{m} P_{ij}\alpha_j \\ &= \sum_{j=1}^{m} \sum_{i=1}^{r} b_{k_i}P_{ij}\alpha_j\end{aligned}$$

and thus

$$x_j = \sum_{i=1}^{r} b_{k_i}P_{ij}$$

is one possible choice for the x_j (there may be many).

The question of whether $\beta = (b_1, \ldots, b_n)$ is a linear combination of the α_i, and if so, what the scalars x_i are, can also be looked at by asking whether the system of equations

$$\sum_{i=1}^{m} A_{ij}x_i = b_j, \qquad j = 1, \ldots, n$$

has a solution and what the solutions are. The coefficient matrix of this system of equations is the $n \times m$ matrix B with *column* vectors $\alpha_1, \ldots, \alpha_m$. In Chapter 1 we discussed the use of elementary row operations in solving a system of equations $BX = Y$. Let us consider one example in which we adopt both points of view in answering questions about subspaces of F^n.

EXAMPLE 21. Let us pose the following problem. Let W be the subspace of R^4 spanned by the vectors

$$\alpha_1 = (1, 2, 2, 1)$$
$$\alpha_2 = (0, 2, 0, 1)$$
$$\alpha_3 = (-2, 0, -4, 3).$$

(a) Prove that α_1, α_2, α_3 form a basis for W, i.e., that these vectors are linearly independent.

(b) Let $\beta = (b_1, b_2, b_3, b_4)$ be a vector in W. What are the coordinates of β relative to the ordered basis $\{\alpha_1, \alpha_2, \alpha_3\}$?

(c) Let

$$\alpha_1' = (1, 0, 2, 0)$$
$$\alpha_2' = (0, 2, 0, 1)$$
$$\alpha_3' = (0, 0, 0, 3).$$

Show that α_1', α_2', α_3' form a basis for W.

(d) If β is in W, let X denote the coordinate matrix of β relative to the α-basis and X' the coordinate matrix of β relative to the α'-basis. Find the 3×3 matrix P such that $X = PX'$ for every such β.

To answer these questions by the first method we form the matrix A with row vectors α_1, α_2, α_3, find the row-reduced echelon matrix R which is row-equivalent to A and simultaneously perform the same operations on the identity to obtain the invertible matrix Q such that $R = QA$:

$$\begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix} \to R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \to Q = \tfrac{1}{6}\begin{bmatrix} 6 & -6 & 0 \\ -2 & 5 & -1 \\ 4 & -4 & 2 \end{bmatrix}$$

(a) Clearly R has rank 3, so α_1, α_2 and α_3 are independent.

(b) Which vectors $\beta = (b_1, b_2, b_3, b_4)$ are in W? We have the basis for W given by ρ_1, ρ_2, ρ_3, the row vectors of R. One can see at a glance that the span of ρ_1, ρ_2, ρ_3 consists of the vectors β for which $b_3 = 2b_1$. For such a β we have

$$\begin{aligned} \beta &= b_1\rho_1 + b_2\rho_2 + b_4\rho_3 \\ &= [b_1, b_2, b_4]R \\ &= [b_1 \quad b_2 \quad b_4]QA \\ &= x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3 \end{aligned}$$

where $x_i = [b_1 \quad b_2 \quad b_4]Q_i$:

$$\begin{aligned} x_1 &= \quad b_1 - \tfrac{1}{3}b_2 + \tfrac{2}{3}b_4 \\ x_2 &= -b_1 + \tfrac{5}{6}b_2 - \tfrac{2}{3}b_4 \\ x_3 &= \quad - \tfrac{1}{6}b_2 + \tfrac{1}{3}b_4. \end{aligned} \tag{2-26}$$

(c) The vectors α_1', α_2', α_3' are all of the form (y_1, y_2, y_3, y_4) with $y_3 = 2y_1$ and thus they are in W. One can see at a glance that they are independent.

(d) The matrix P has for its columns

$$P_j = [\alpha_j']_{\mathcal{B}}$$

where $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$. The equations (2-26) tell us how to find the coordinate matrices for α_1', α_2', α_3'. For example with $\beta = \alpha_1'$ we have $b_1 = 1$, $b_2 = 0$, $b_3 = 2$, $b_4 = 0$, and

$$\begin{aligned} x_1 &= \quad 1 - \tfrac{1}{3}(0) + \tfrac{2}{3}(0) = \quad 1 \\ x_2 &= -1 + \tfrac{5}{6}(0) - \tfrac{2}{3}(0) = -1 \\ x_3 &= \quad - \tfrac{1}{6}(0) + \tfrac{1}{3}(0) = \quad 0. \end{aligned}$$

Thus $\alpha_1' = \alpha_1 - \alpha_2$. Similarly we obtain $\alpha_2' = \alpha_2$ and $\alpha_3' = 2\alpha_1 - 2\alpha_2 + \alpha_3$. Hence

$$P = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

Now let us see how we would answer the questions by the second method which we described. We form the 4×3 matrix B with column vectors α_1, α_2, α_3:

$$B = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 2 & 0 \\ 2 & 0 & -4 \\ 1 & 1 & 3 \end{bmatrix}$$

We inquire for which y_1, y_2, y_3, y_4 the system $BX = Y$ has a solution.

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 2 & 2 & 0 & y_2 \\ 2 & 0 & -4 & y_3 \\ 1 & 1 & 3 & y_4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 2 & 4 & y_2 - 2y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \\ 0 & 1 & 5 & y_4 - y_1 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 0 & -6 & y_2 - 2y_4 \\ 0 & 1 & 5 & y_4 - y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & y_1 - \frac{1}{3}y_2 + \frac{2}{3}y_4 \\ 0 & 0 & 1 & \frac{1}{6}(2y_4 - y_2) \\ 0 & 1 & 0 & -y_1 + \frac{5}{6}y_2 - \frac{2}{3}y_4 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix}$$

Thus the condition that the system $BX = Y$ have a solution is $y_3 = 2y_1$. So $\beta = (b_1, b_2, b_3, b_4)$ is in W if and only if $b_3 = 2b_1$. If β is in W, then the coordinates (x_1, x_2, x_3) in the ordered basis $\{\alpha_1, \alpha_2, \alpha_3\}$ can be read off from the last matrix above. We obtain once again the formulas (2-26) for those coordinates.

The questions (c) and (d) are now answered as before.

EXAMPLE 22. We consider the 5×5 matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and the following problems concerning A

(a) Find an invertible matrix P such that PA is a row-reduced echelon matrix R.

(b) Find a basis for the row space W of A.

(c) Say which vectors $(b_1, b_2, b_3, b_4, b_5)$ are in W.

(d) Find the coordinate matrix of each vector $(b_1, b_2, b_3, b_4, b_5)$ in W in the ordered basis chosen in (b).

(e) Write each vector $(b_1, b_2, b_3, b_4, b_5)$ in W as a linear combination of the rows of A.

(f) Give an explicit description of the vector space V of all 5×1 column matrices X such that $AX = 0$.

(g) Find a basis for V.

(h) For what 5×1 column matrices Y does the equation $AX = Y$ have solutions X?

To solve these problems we form the augmented matrix A' of the system $AX = Y$ and apply an appropriate sequence of row operations to A'.

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 1 & 2 & -1 & -1 & 0 & y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 2 & 4 & 1 & 10 & 1 & y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & -1 & -4 & 0 & -y_1 + y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 0 & 0 & 1 & 4 & 1 & -2y_1 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 1 & -3y_1 + y_2 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 1 & y_5 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 0 & -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

(a) If

$$PY = \begin{bmatrix} y_1 \\ y_1 - y_2 \\ y_5 \\ -y_1 + y_2 + y_3 \\ -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

for all Y, then

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix}$$

hence PA is the row-reduced echelon matrix

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It should be stressed that the matrix P is not unique. There are, in fact, many invertible matrices P (which arise from different choices for the operations used to reduce A') such that $PA = R$.

(b) As a basis for W we may take the non-zero rows

$$\begin{aligned} \rho_1 &= (1 \quad 2 \quad 0 \quad 3 \quad 0) \\ \rho_2 &= (0 \quad 0 \quad 1 \quad 4 \quad 0) \\ \rho_3 &= (0 \quad 0 \quad 0 \quad 0 \quad 1) \end{aligned}$$

of R.

(c) The row-space W consists of all vectors of the form

$$\begin{aligned} \beta &= c_1\rho_1 + c_2\rho_2 + c_3\rho_3 \\ &= (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3) \end{aligned}$$

where c_1, c_2, c_3 are arbitrary scalars. Thus $(b_1, b_2, b_3, b_4, b_5)$ is in W if and only if

$$(b_1, b_2, b_3, b_4, b_5) = b_1\rho_1 + b_3\rho_2 + b_5\rho_3$$

which is true if and only if

$$\begin{aligned} b_2 &= 2b_1 \\ b_4 &= 3b_1 + 4b_3. \end{aligned}$$

These equations are instances of the general system (2-25), and using them we may tell at a glance whether a given vector lies in W. Thus $(-5, -10, 1, -11, 20)$ is a linear combination of the rows of A, but $(1, 2, 3, 4, 5)$ is not.

(d) The coordinate matrix of the vector $(b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5)$ in the basis $\{\rho_1, \rho_2, \rho_3\}$ is evidently

$$\begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}.$$

(e) There are many ways to write the vectors in W as linear combinations of the rows of A. Perhaps the easiest method is to follow the first procedure indicated before Example 21:

$$\begin{aligned} \beta &= (b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5) \\ &= [b_1, b_3, b_5, 0, 0] \cdot R \\ &= [b_1, b_3, b_5, 0, 0] \cdot PA \\ &= [b_1, b_3, b_5, 0, 0] \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix} \cdot A \\ &= [b_1 + b_3, -b_3, 0, 0, b_5] \cdot A. \end{aligned}$$

In particular, with $\beta = (-5, -10, 1, -11, 20)$ we have

$$\beta = (-4, -1, 0, 0, 20) \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(f) The equations in the system $RX = 0$ are

$$\begin{aligned} x_1 + 2x_2 + 3x_4 &= 0 \\ x_3 + 4x_4 &= 0 \\ x_5 &= 0. \end{aligned}$$

Thus V consists of all columns of the form

$$X = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \\ 0 \end{bmatrix}$$

where x_2 and x_4 are arbitrary.

(g) The columns

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}$$

form a basis of V. This is an example of the basis described in Example 15.

(h) The equation $AX = Y$ has solutions X if and only if

$$\begin{aligned} -y_1 + y_2 + y_3 &= 0 \\ -3y_1 + y_2 + y_4 - y_5 &= 0. \end{aligned}$$

Exercises

1. Let $s < n$ and A an $s \times n$ matrix with entries in the field F. Use Theorem 4 (not its proof) to show that there is a non-zero X in $F^{n\times 1}$ such that $AX = 0$.

2. Let

$$\alpha_1 = (1, 1, -2, 1), \quad \alpha_2 = (3, 0, 4, -1), \quad \alpha_3 = (-1, 2, 5, 2).$$

Let

$$\alpha = (4, -5, 9, -7), \quad \beta = (3, 1, -4, 4), \quad \gamma = (-1, 1, 0, 1).$$

(a) Which of the vectors α, β, γ are in the subspace of R^4 spanned by the α_i?
(b) Which of the vectors α, β, γ are in the subspace of C^4 spanned by the α_i?
(c) Does this suggest a theorem?

3. Consider the vectors in R^4 defined by

$$\alpha_1 = (-1, 0, 1, 2), \quad \alpha_2 = (3, 4, -2, 5), \quad \alpha_3 = (1, 4, 0, 9).$$

Find a system of homogeneous linear equations for which the space of solutions is exactly the subspace of R^4 spanned by the three given vectors.

4. In C^3, let

$$\alpha_1 = (1, 0, -i), \quad \alpha_2 = (1 + i, 1 - i, 1), \quad \alpha_3 = (i, i, i).$$

Prove that these vectors form a basis for C^3. What are the coordinates of the vector (a, b, c) in this basis?

5. Give an explicit description of the type (2-25) for the vectors

$$\beta = (b_1, b_2, b_3, b_4, b_5)$$

in R^5 which are linear combinations of the vectors

$$\begin{aligned} \alpha_1 &= (1, 0, 2, 1, -1), \quad & \alpha_2 &= (-1, 2, -4, 2, 0) \\ \alpha_3 &= (2, -1, 5, 2, 1), \quad & \alpha_4 &= (2, 1, 3, 5, 2). \end{aligned}$$

6. Let V be the real vector space spanned by the rows of the matrix

$$A = \begin{bmatrix} 3 & 21 & 0 & 9 & 0 \\ 1 & 7 & -1 & -2 & -1 \\ 2 & 14 & 0 & 6 & 1 \\ 6 & 42 & -1 & 13 & 0 \end{bmatrix}.$$

(a) Find a basis for V.
(b) Tell which vectors $(x_1, x_2, x_3, x_4, x_5)$ are elements of V.
(c) If $(x_1, x_2, x_3, x_4, x_5)$ is in V what are its coordinates in the basis chosen in part (a)?

7. Let A be an $m \times n$ matrix over the field F, and consider the system of equations $AX = Y$. Prove that this system of equations has a solution if and only if the row rank of A is equal to the row rank of the augmented matrix of the system.

3. Linear Transformations

3.1. Linear Transformations

We shall now introduce linear transformations, the objects which we shall study in most of the remainder of this book. The reader may find it helpful to read (or reread) the discussion of functions in the Appendix, since we shall freely use the terminology of that discussion.

Definition. *Let* V *and* W *be vector spaces over the field* F. *A* **linear transformation from** V **into** W *is a function* T *from* V *into* W *such that*

$$T(c\alpha + \beta) = c(T\alpha) + T\beta$$

for all α *and* β *in* V *and all scalars* c *in* F.

EXAMPLE 1. If V is any vector space, the identity transformation I, defined by $I\alpha = \alpha$, is a linear transformation from V into V. The **zero transformation** 0, defined by $0\alpha = 0$, is a linear transformation from V into V.

EXAMPLE 2. Let F be a field and let V be the space of polynomial functions f from F into F, given by

$$f(x) = c_0 + c_1x + \cdots + c_kx^k.$$

Let

$$(Df)(x) = c_1 + 2c_2x + \cdots + kc_kx^{k-1}.$$

Then D is a linear transformation from V into V—the differentiation transformation.

EXAMPLE 3. Let A be a fixed $m \times n$ matrix with entries in the field F. The function T defined by $T(X) = AX$ is a linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$. The function U defined by $U(\alpha) = \alpha A$ is a linear transformation from F^m into F^n.

EXAMPLE 4. Let P be a fixed $m \times m$ matrix with entries in the field F and let Q be a fixed $n \times n$ matrix over F. Define a function T from the space $F^{m\times n}$ into itself by $T(A) = PAQ$. Then T is a linear transformation from $F^{m\times n}$ into $F^{m\times n}$, because

$$\begin{aligned} T(cA + B) &= P(cA + B)Q \\ &= (cPA + PB)Q \\ &= cPAQ + PBQ \\ &= cT(A) + T(B). \end{aligned}$$

EXAMPLE 5. Let R be the field of real numbers and let V be the space of all functions from R into R which are *continuous*. Define T by

$$(Tf)(x) = \int_0^x f(t)\, dt.$$

Then T is a linear transformation from V into V. The function Tf is not only continuous but has a continuous first derivative. The linearity of integration is one of its fundamental properties.

The reader should have no difficulty in verifying that the transformations defined in Examples 1, 2, 3, and 5 are linear transformations. We shall expand our list of examples considerably as we learn more about linear transformations.

It is important to note that if T is a linear transformation from V into W, then $T(0) = 0$; one can see this from the definition because

$$T(0) = T(0 + 0) = T(0) + T(0).$$

This point is often confusing to the person who is studying linear algebra for the first time, since he probably has been exposed to a slightly different use of the term 'linear function.' A brief comment should clear up the confusion. Suppose V is the vector space R^1. A linear transformation from V into V is then a particular type of real-valued function on the real line R. In a calculus course, one would probably call such a function linear if its graph is a straight line. A linear transformation from R^1 into R^1, according to our definition, will be a function from R into R, the graph of which is a straight line *passing through the origin.*

In addition to the property $T(0) = 0$, let us point out another property of the general linear transformation T. Such a transformation 'preserves' linear combinations; that is, if $\alpha_1, \ldots, \alpha_n$ are vectors in V and $c_1, \ldots, c_n$ are scalars, then

$$T(c_1\alpha_1 + \cdots + c_n\alpha_n) = c_1(T\alpha_1) + \cdots + c_n(T\alpha_n).$$

This follows readily from the definition. For example,

$$\begin{aligned} T(c_1\alpha_1 + c_2\alpha_2) &= c_1(T\alpha_1) + T(c_2\alpha_2) \\ &= c_1(T\alpha_1) + c_2(T\alpha_2). \end{aligned}$$

Theorem 1. *Let* V *be a finite-dimensional vector space over the field* F *and let* $\{\alpha_1, \ldots, \alpha_n\}$ *be an ordered basis for* V. *Let* W *be a vector space over the same field* F *and let* $\beta_1, \ldots, \beta_n$ *be any vectors in* W. *Then there is precisely one linear transformation* T *from* V *into* W *such that*

$$\mathrm{T}\alpha_\mathrm{j} = \beta_\mathrm{j}, \qquad \mathrm{j} = 1, \ldots, \mathrm{n}.$$

Proof. To prove there is some linear transformation T with $T\alpha_j = \beta_j$ we proceed as follows. Given α in V, there is a unique n-tuple $(x_1, \ldots, x_n)$ such that

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n.$$

For this vector α we define

$$T\alpha = x_1\beta_1 + \cdots + x_n\beta_n.$$

Then T is a well-defined rule for associating with each vector α in V a vector $T\alpha$ in W. From the definition it is clear that $T\alpha_j = \beta_j$ for each j. To see that T is linear, let

$$\beta = y_1\alpha_1 + \cdots + y_n\alpha_n$$

be in V and let c be any scalar. Now

$$c\alpha + \beta = (cx_1 + y_1)\alpha_1 + \cdots + (cx_n + y_n)\alpha_n$$

and so by definition

$$T(c\alpha + \beta) = (cx_1 + y_1)\beta_1 + \cdots + (cx_n + y_n)\beta_n.$$

On the other hand,

$$\begin{aligned} c(T\alpha) + T\beta &= c\sum_{i=1}^{n} x_i\beta_i + \sum_{i=1}^{n} y_i\beta_i \\ &= \sum_{i=1}^{n} (cx_i + y_i)\beta_i \end{aligned}$$

and thus

$$T(c\alpha + \beta) = c(T\alpha) + T\beta.$$

If U is a linear transformation from V into W with $U\alpha_j = \beta_j$, $j = 1, \ldots, n$, then for the vector $\alpha = \sum_{i=1}^{n} x_i\alpha_i$ we have

$$\begin{aligned} U\alpha &= U\left(\sum_{i=1}^{n} x_i\alpha_i\right) \\ &= \sum_{i=1}^{n} x_i(U\alpha_i) \\ &= \sum_{i=1}^{n} x_i\beta_i \end{aligned}$$

so that U is exactly the rule T which we defined above. This shows that the linear transformation T with $T\alpha_j = \beta_j$ is unique. ∎

Theorem 1 is quite elementary; however, it is so basic that we have stated it formally. The concept of function is very general. If V and W are (non-zero) vector spaces, there is a multitude of functions from V into W. Theorem 1 helps to underscore the fact that the functions which are linear are extremely special.

EXAMPLE 6. The vectors

$$\alpha_1 = (1, 2)$$
$$\alpha_2 = (3, 4)$$

are linearly independent and therefore form a basis for R^2. According to Theorem 1, there is a unique linear transformation from R^2 into R^3 such that

$$T\alpha_1 = (3, 2, 1)$$
$$T\alpha_2 = (6, 5, 4).$$

If so, we must be able to find $T(\epsilon_1)$. We find scalars c_1, c_2 such that $\epsilon_1 = c_1\alpha_1 + c_2\alpha_2$ and then we know that $T\epsilon_1 = c_1T\alpha_1 + c_2T\alpha_2$. If $(1, 0) = c_1(1, 2) + c_2(3, 4)$ then $c_1 = -2$ and $c_2 = 1$. Thus

$$\begin{aligned} T(1, 0) &= -2(3, 2, 1) + (6, 5, 4) \\ &= (0, 1, 2). \end{aligned}$$

EXAMPLE 7. Let T be a linear transformation from the m-tuple space F^m into the n-tuple space F^n. Theorem 1 tells us that T is uniquely determined by the sequence of vectors $\beta_1, \ldots, \beta_m$ where

$$\beta_i = T\epsilon_i, \qquad i = 1, \ldots, m.$$

In short, T is uniquely determined by the images of the standard basis vectors. The determination is

$$\alpha = (x_1, \ldots, x_m)$$
$$T\alpha = x_1\beta_1 + \cdots + x_m\beta_m.$$

If B is the $m \times n$ matrix which has row vectors $\beta_1, \ldots, \beta_m$, this says that

$$T\alpha = \alpha B.$$

In other words, if $\beta_i = (B_{i1}, \ldots, B_{in})$, then

$$T(x_1, \ldots, x_n) = [x_1 \cdots x_m] \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{m1} & \cdots & B_{mn} \end{bmatrix}.$$

This is a very explicit description of the linear transformation. In Section 3.4 we shall make a serious study of the relationship between linear trans-

formations and matrices. We shall not pursue the particular description $T\alpha = \alpha B$ because it has the matrix B on the right of the vector α, and that can lead to some confusion. The point of this example is to show that we can give an explicit and reasonably simple description of all linear transformations from F^m into F^n.

If T is a linear transformation from V into W, then the range of T is not only a subset of W; it is a subspace of W. Let R_T be the range of T, that is, the set of all vectors β in W such that $\beta = T\alpha$ for some α in V. Let β_1 and β_2 be in R_T and let c be a scalar. There are vectors α_1 and α_2 in V such that $T\alpha_1 = \beta_1$ and $T\alpha_2 = \beta_2$. Since T is linear

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c\beta_1 + \beta_2, \end{aligned}$$

which shows that $c\beta_1 + \beta_2$ is also in R_T.

Another interesting subspace associated with the linear transformation T is the set N consisting of the vectors α in V such that $T\alpha = 0$. It is a subspace of V because

(a) $T(0) = 0$, so that N is non-empty;
(b) if $T\alpha_1 = T\alpha_2 = 0$, then

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c0 + 0 \\ &= 0 \end{aligned}$$

so that $c\alpha_1 + \alpha_2$ is in N.

Definition. *Let* V *and* W *be vector spaces over the field* F *and let* T *be a linear transformation from* V *into* W. *The* **null space** *of* T *is the set of all vectors* α *in* V *such that* $T\alpha = 0$.

If V *is finite-dimensional, the* **rank** *of* T *is the dimension of the range of* T *and the* **nullity** *of* T *is the dimension of the null space of* T.

The following is one of the most important results in linear algebra.

Theorem 2. *Let* V *and* W *be vector spaces over the field* F *and let* T *be a linear transformation from* V *into* W. *Suppose that* V *is finite-dimensional. Then*

$$\textit{rank}\ (\mathrm{T}) + \textit{nullity}\ (\mathrm{T}) = \textit{dim}\ \mathrm{V}.$$

Proof. Let $\{\alpha_1, \ldots, \alpha_k\}$ be a basis for N, the null space of T. There are vectors $\alpha_{k+1}, \ldots, \alpha_n$ in V such that $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V. We shall now prove that $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ is a basis for the range of T. The vectors $T\alpha_1, \ldots, T\alpha_n$ certainly span the range of T, and since $T\alpha_j = 0$, for $j \leq k$, we see that $T\alpha_{k+1}, \ldots, T\alpha_n$ span the range. To see that these vectors are independent, suppose we have scalars c_i such that

$$\sum_{i=k+1}^{n} c_i(T\alpha_i) = 0.$$

This says that

$$T\left(\sum_{i=k+1}^{n} c_i\alpha_i\right) = 0$$

and accordingly the vector $\alpha = \sum_{i=k+1}^{n} c_i\alpha_i$ is in the null space of T. Since $\alpha_1, \ldots, \alpha_k$ form a basis for N, there must be scalars $b_1, \ldots, b_k$ such that

$$\alpha = \sum_{i=1}^{k} b_i\alpha_i.$$

Thus

$$\sum_{i=1}^{k} b_i\alpha_i - \sum_{j=k+1}^{n} c_j\alpha_j = 0$$

and since $\alpha_1, \ldots, \alpha_n$ are linearly independent we must have

$$b_1 = \cdots = b_k = c_{k+1} = \cdots = c_n = 0.$$

If r is the rank of T, the fact that $T\alpha_{k+1}, \ldots, T\alpha_n$ form a basis for the range of T tells us that $r = n - k$. Since k is the nullity of T and n is the dimension of V, we are done. ∎

Theorem 3. *If* A *is an* m × n *matrix with entries in the field* F, *then*

$$\textit{row rank } (A) = \textit{column rank } (A).$$

Proof. Let T be the linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$ defined by $T(X) = AX$. The null space of T is the solution space for the system $AX = 0$, i.e., the set of all column matrices X such that $AX = 0$. The range of T is the set of all $m \times 1$ column matrices Y such that $AX = Y$ has a solution for X. If $A_1, \ldots, A_n$ are the columns of A, then

$$AX = x_1A_1 + \cdots + x_nA_n$$

so that the range of T is the subspace spanned by the columns of A. In other words, the range of T is the column space of A. Therefore,

$$\text{rank } (T) = \text{column rank } (A).$$

Theorem 2 tells us that if S is the solution space for the system $AX = 0$, then

$$\dim S + \text{column rank } (A) = n.$$

We now refer to Example 15 of Chapter 2. Our deliberations there showed that, if r is the dimension of the row space of A, then the solution space S has a basis consisting of $n - r$ vectors:

$$\dim S = n - \text{row rank } (A).$$

It is now apparent that

$$\text{row rank } (A) = \text{column rank } (A). \quad ∎$$

The proof of Theorem 3 which we have just given depends upon

explicit calculations concerning systems of linear equations. There is a more conceptual proof which does not rely on such calculations. We shall give such a proof in Section 3.7.

Exercises

1. Which of the following functions T from R^2 into R^2 are linear transformations?

(a) $T(x_1, x_2) = (1 + x_1, x_2)$;
(b) $T(x_1, x_2) = (x_2, x_1)$;
(c) $T(x_1, x_2) = (x_1^2, x_2)$;
(d) $T(x_1, x_2) = (\sin x_1, x_2)$;
(e) $T(x_1, x_2) = (x_1 - x_2, 0)$.

2. Find the range, rank, null space, and nullity for the zero transformation and the identity transformation on a finite-dimensional space V.

3. Describe the range and the null space for the differentiation transformation of Example 2. Do the same for the integration transformation of Example 5.

4. Is there a linear transformation T from R^3 into R^2 such that $T(1, -1, 1) = (1, 0)$ and $T(1, 1, 1) = (0, 1)$?

5. If

$$\begin{aligned} \alpha_1 &= (1, -1), & \beta_1 &= (1, 0) \\ \alpha_2 &= (2, -1), & \beta_2 &= (0, 1) \\ \alpha_3 &= (-3, 2), & \beta_3 &= (1, 1) \end{aligned}$$

is there a linear transformation T from R^2 into R^2 such that $T\alpha_i = \beta_i$ for $i = 1, 2$ and 3?

6. Describe explicitly (as in Exercises 1 and 2) the linear transformation T from F^2 into F^2 such that $T\epsilon_1 = (a, b)$, $T\epsilon_2 = (c, d)$.

7. Let F be a subfield of the complex numbers and let T be the function from F^3 into F^3 defined by

$$T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 2x_1 + x_2, -x_1 - 2x_2 + 2x_3).$$

(a) Verify that T is a linear transformation.

(b) If (a, b, c) is a vector in F^3, what are the conditions on a, b, and c that the vector be in the range of T? What is the rank of T?

(c) What are the conditions on a, b, and c that (a, b, c) be in the null space of T? What is the nullity of T?

8. Describe explicitly a linear transformation from R^3 into R^3 which has as its range the subspace spanned by $(1, 0, -1)$ and $(1, 2, 2)$.

9. Let V be the vector space of all $n \times n$ matrices over the field F, and let B be a fixed $n \times n$ matrix. If

$$T(A) = AB - BA$$

verify that T is a linear transformation from V into V.

10. Let V be the set of all complex numbers regarded as a vector space over the

field of *real* numbers (usual operations). Find a function from V into V which is a linear transformation on the above vector space, but which is not a linear transformation on C^1, i.e., which is not complex linear.

11. Let V be the space of $n \times 1$ matrices over F and let W be the space of $m \times 1$ matrices over F. Let A be a fixed $m \times n$ matrix over F and let T be the linear transformation from V into W defined by $T(X) = AX$. Prove that T is the zero transformation if and only if A is the zero matrix.

12. Let V be an n-dimensional vector space over the field F and let T be a linear transformation from V into V such that the range and null space of T are identical. Prove that n is even. (Can you give an example of such a linear transformation T?)

13. Let V be a vector space and T a linear transformation from V into V. Prove that the following two statements about T are equivalent.

(a) The intersection of the range of T and the null space of T is the zero subspace of V.

(b) If $T(T\alpha) = 0$, then $T\alpha = 0$.

3.2. The Algebra of Linear Transformations

In the study of linear transformations from V into W, it is of fundamental importance that the set of these transformations inherits a natural vector space structure. The set of linear transformations from a space V into itself has even more algebraic structure, because ordinary composition of functions provides a 'multiplication' of such transformations. We shall explore these ideas in this section.

Theorem 4. *Let* V *and* W *be vector spaces over the field* F. *Let* T *and* U *be linear transformations from* V *into* W. *The function* (T + U) *defined by*

$$(\mathrm{T} + \mathrm{U})(\alpha) = \mathrm{T}\alpha + \mathrm{U}\alpha$$

is a linear transformation from V *into* W. *If* c *is any element of* F, *the function* (cT) *defined by*

$$(\mathrm{cT})(\alpha) = \mathrm{c}(\mathrm{T}\alpha)$$

is a linear transformation from V *into* W. *The set of all linear transformations from* V *into* W, *together with the addition and scalar multiplication defined above, is a vector space over the field* F.

Proof. Suppose T and U are linear transformations from V into W and that we define $(T + U)$ as above. Then

$$\begin{aligned}(T + U)(c\alpha + \beta) &= T(c\alpha + \beta) + U(c\alpha + \beta)\\ &= c(T\alpha) + T\beta + c(U\alpha) + U\beta\\ &= c(T\alpha + U\alpha) + (T\beta + U\beta)\\ &= c(T + U)(\alpha) + (T + U)(\beta)\end{aligned}$$

which shows that $(T + U)$ is a linear transformation. Similarly,

$$\begin{aligned}(cT)(d\alpha + \beta) &= c[T(d\alpha + \beta)] \\ &= c[d(T\alpha) + T\beta] \\ &= cd(T\alpha) + c(T\beta) \\ &= d[c(T\alpha)] + c(T\beta) \\ &= d[(cT)\alpha] + (cT)\beta\end{aligned}$$

which shows that (cT) is a linear transformation.

To verify that the set of linear transformations of V into W (together with these operations) is a vector space, one must directly check each of the conditions on the vector addition and scalar multiplication. We leave the bulk of this to the reader, and content ourselves with this comment: The zero vector in this space will be the zero transformation, which sends every vector of V into the zero vector in W; each of the properties of the two operations follows from the corresponding property of the operations in the space W. ∎

We should perhaps mention another way of looking at this theorem. If one defines sum and scalar multiple as we did above, then the set of *all* functions from V into W becomes a vector space over the field F. This has nothing to do with the fact that V is a vector space, only that V is a non-empty set. When V is a vector space we can define a linear transformation from V into W, and Theorem 4 says that the linear transformations are a subspace of the space of all functions from V into W.

We shall denote the space of linear transformations from V into W by $L(V, W)$. We remind the reader that $L(V, W)$ is defined only when V and W are vector spaces over the same field.

Theorem 5. *Let* V *be an* n*-dimensional vector space over the field* F*, and let* W *be an* m*-dimensional vector space over* F*. Then the space* L(V, W) *is finite-dimensional and has dimension* mn.

Proof. Let

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{and} \quad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$$

be ordered bases for V and W, respectively. For each pair of integers (p, q) with $1 \leq p \leq m$ and $1 \leq q \leq n$, we define a linear transformation $E^{p,q}$ from V into W by

$$\begin{aligned}E^{p,q}(\alpha_i) &= \begin{cases} 0, & \text{if } i \neq q \\ \beta_p, & \text{if } i = q \end{cases} \\ &= \delta_{iq}\beta_p.\end{aligned}$$

According to Theorem 1, there is a unique linear transformation from V into W satisfying these conditions. The claim is that the mn transformations $E^{p,q}$ form a basis for $L(V, W)$.

Let T be a linear transformation from V into W. For each j, $1 \leq j \leq n$,

let $A_{ij}, \ldots, A_{mj}$ be the coordinates of the vector $T\alpha_j$ in the ordered basis $\mathcal{B}'$, i.e.,

$$T\alpha_j = \sum_{p=1}^{m} A_{pj}\beta_p. \tag{3-1}$$

We wish to show that

$$T = \sum_{p=1}^{m}\sum_{q=1}^{n} A_{pq}E^{p,q}. \tag{3-2}$$

Let U be the linear transformation in the right-hand member of (3-2). Then for each j

$$\begin{aligned} U\alpha_j &= \sum_p \sum_q A_{pq}E^{p,q}(\alpha_j) \\ &= \sum_p \sum_q A_{pq}\delta_{jq}\beta_p \\ &= \sum_{p=1}^{m} A_{pj}\beta_p \\ &= T\alpha_j \end{aligned}$$

and consequently $U = T$. Now (3-2) shows that the $E^{p,q}$ span $L(V, W)$; we must prove that they are independent. But this is clear from what we did above; for, if the transformation

$$U = \sum_p \sum_q A_{pq}E^{p,q}$$

is the zero transformation, then $U\alpha_j = 0$ for each j, so

$$\sum_{p=1}^{m} A_{pj}\beta_p = 0$$

and the independence of the β_p implies that $A_{pj} = 0$ for every p and j. ∎

Theorem 6. *Let* V, W, *and* Z *be vector spaces over the field* F. *Let* T *be a linear transformation from* V *into* W *and* U *a linear transformation from* W *into* Z. *Then the composed function* UT *defined by* $(\mathrm{UT})(\alpha) = \mathrm{U}(\mathrm{T}(\alpha))$ *is a linear transformation from* V *into* Z.

Proof.

$$\begin{aligned} (UT)(c\alpha + \beta) &= U[T(c\alpha + \beta)] \\ &= U(cT\alpha + T\beta) \\ &= c[U(T\alpha)] + U(T\beta) \\ &= c(UT)(\alpha) + (UT)(\beta). \end{aligned}$$ ∎

In what follows, we shall be primarily concerned with linear transformation of a vector space into itself. Since we would so often have to write 'T is a linear transformation from V into V,' we shall replace this with 'T is a linear operator on V.'

Definition. *If* V *is a vector space over the field* F, *a* **linear operator on** V *is a linear transformation from* V *into* V.

In the case of Theorem 6 when $V = W = Z$, so that U and T are linear operators on the space V, we see that the composition UT is again a linear operator on V. Thus the space $L(V, V)$ has a 'multiplication' defined on it by composition. In this case the operator TU is also defined, and one should note that in general $UT \neq TU$, i.e., $UT - TU \neq 0$. We should take special note of the fact that if T is a linear operator on V then we can compose T with T. We shall use the notation $T^2 = TT$, and in general $T^n = T \cdots T$ (n times) for $n = 1, 2, 3, \ldots$. We define $T^0 = I$ if $T \neq 0$.

Lemma. *Let* V *be a vector space over the field* F; *let* U, T_1 *and* T_2 *be linear operators on* V; *let* c *be an element of* F.

(a) $IU = UI = U$;

(b) $U(T_1 + T_2) = UT_1 + UT_2$; $(T_1 + T_2)U = T_1U + T_2U$;

(c) $c(UT_1) = (cU)T_1 = U(cT_1)$.

Proof. (a) This property of the identity function is obvious. We have stated it here merely for emphasis.

(b)
$$\begin{aligned}[U(T_1 + T_2)](\alpha) &= U[(T_1 + T_2)(\alpha)] \\ &= U(T_1\alpha + T_2\alpha) \\ &= U(T_1\alpha) + U(T_2\alpha) \\ &= (UT_1)(\alpha) + (UT_2)(\alpha)\end{aligned}$$

so that $U(T_1 + T_2) = UT_1 + UT_2$. Also

$$\begin{aligned}[(T_1 + T_2)U](\alpha) &= (T_1 + T_2)(U\alpha) \\ &= T_1(U\alpha) + T_2(U\alpha) \\ &= (T_1U)(\alpha) + (T_2U)(\alpha)\end{aligned}$$

so that $(T_1 + T_2)U = T_1U + T_2U$. (The reader may note that the proofs of these two distributive laws do not use the fact that T_1 and T_2 are linear, and the proof of the second one does not use the fact that U is linear either.)

(c) We leave the proof of part (c) to the reader. ∎

The contents of this lemma and a portion of Theorem 5 tell us that the vector space $L(V, V)$, together with the composition operation, is what is known as a linear algebra with identity. We shall discuss this in Chapter 4.

EXAMPLE 8. If A is an $m \times n$ matrix with entries in F, we have the linear transformation T defined by $T(X) = AX$, from $F^{n\times 1}$ into $F^{m\times 1}$. If B is a $p \times m$ matrix, we have the linear transformation U from $F^{m\times 1}$ into $F^{p\times 1}$ defined by $U(Y) = BY$. The composition UT is easily described:

$$\begin{aligned}(UT)(X) &= U(T(X)) \\ &= U(AX) \\ &= B(AX) \\ &= (BA)X.\end{aligned}$$

Thus UT is 'left multiplication by the product matrix BA.'

EXAMPLE 9. Let F be a field and V the vector space of all polynomial functions from F into F. Let D be the differentiation operator defined in Example 2, and let T be the linear operator 'multiplication by x':

$$(Tf)(x) = xf(x).$$

Then $DT \neq TD$. In fact, the reader should find it easy to verify that $DT - TD = I$, the identity operator.

Even though the 'multiplication' we have on $L(V, V)$ is not commutative, it is nicely related to the vector space operations of $L(V, V)$.

EXAMPLE 10. Let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ be an ordered basis for a vector space V. Consider the linear operators $E^{p,q}$ which arose in the proof of Theorem 5:

$$E^{p,q}(\alpha_i) = \delta_{iq}\alpha_p.$$

These n^2 linear operators form a basis for the space of linear operators on V. What is $E^{p,q}E^{r,s}$? We have

$$\begin{aligned}(E^{p,q}E^{r,s})(\alpha_i) &= E^{p,q}(\delta_{is}\alpha_r) \\ &= \delta_{is}E^{p,q}(\alpha_r) \\ &= \delta_{is}\delta_{rq}\alpha_p.\end{aligned}$$

Therefore,

$$E^{p,q}E^{r,s} = \begin{cases} 0, & \text{if } r \neq q \\ E^{p,s}, & \text{if } q = r. \end{cases}$$

Let T be a linear operator on V. We showed in the proof of Theorem 5 that if

$$\begin{aligned}A_j &= [T\alpha_j]_{\mathcal{B}} \\ A &= [A_1, \ldots, A_n]\end{aligned}$$

then

$$T = \sum_p \sum_q A_{pq}E^{p,q}.$$

If

$$U = \sum_r \sum_s B_{rs}E^{r,s}$$

is another linear operator on V, then the last lemma tells us that

$$\begin{aligned}TU &= (\sum_p \sum_q A_{pq}E^{p,q})(\sum_r \sum_s B_{rs}E^{r,s}) \\ &= \sum_p \sum_q \sum_r \sum_s A_{pq}B_{rs}E^{p,q}E^{r,s}.\end{aligned}$$

As we have noted, the only terms which survive in this huge sum are the terms where $q = r$, and since $E^{p,r}E^{r,s} = E^{p,s}$, we have

$$\begin{aligned}TU &= \sum_p \sum_s (\sum_r A_{pr}B_{rs})E^{p,s} \\ &= \sum_p \sum_s (AB)_{ps}E^{p,s}.\end{aligned}$$

Thus, the effect of composing T and U is to multiply the matrices A and B.

In our discussion of algebraic operations with linear transformations we have not yet said anything about invertibility. One specific question of interest is this. For which linear operators T on the space V does there exist a linear operator T^{-1} such that $TT^{-1} = T^{-1}T = I$?

The function T from V into W is called **invertible** if there exists a function U from W into V such that UT is the identity function on V and TU is the identity function on W. If T is invertible, the function U is unique and is denoted by T^{-1}. (See Appendix.) Furthermore, T is invertible if and only if

1. T is 1:1, that is, $T\alpha = T\beta$ implies $\alpha = \beta$;
2. T is onto, that is, the range of T is (all of) W.

Theorem 7. *Let* V *and* W *be vector spaces over the field* F *and let* T *be a linear transformation from* V *into* W. *If* T *is invertible, then the inverse function* T^{-1} *is a linear transformation from* W *onto* V.

Proof. We repeat ourselves in order to underscore a point. When T is one-one and onto, there is a uniquely determined inverse function T^{-1} which maps W onto V such that $T^{-1}T$ is the identity function on V, and TT^{-1} is the identity function on W. What we are proving here is that if a linear function T is invertible, then the inverse T^{-1} is also linear.

Let β_1 and β_2 be vectors in W and let c be a scalar. We wish to show that

$$T^{-1}(c\beta_1 + \beta_2) = cT^{-1}\beta_1 + T^{-1}\beta_2.$$

Let $\alpha_i = T^{-1}\beta_i$, $i = 1, 2$, that is, let α_i be the unique vector in V such that $T\alpha_i = \beta_i$. Since T is linear,

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c\beta_1 + \beta_2. \end{aligned}$$

Thus $c\alpha_1 + \alpha_2$ is the unique vector in V which is sent by T into $c\beta_1 + \beta_2$, and so

$$\begin{aligned} T^{-1}(c\beta_1 + \beta_2) &= c\alpha_1 + \alpha_2 \\ &= c(T^{-1}\beta_1) + T^{-1}\beta_2 \end{aligned}$$

and T^{-1} is linear. ∎

Suppose that we have an invertible linear transformation T from V onto W and an invertible linear transformation U from W onto Z. Then UT is invertible and $(UT)^{-1} = T^{-1}U^{-1}$. That conclusion does not require the linearity nor does it involve checking separately that UT is 1:1 and onto. All it involves is verifying that $T^{-1}U^{-1}$ is both a left and a right inverse for UT.

If T is linear, then $T(\alpha - \beta) = T\alpha - T\beta$; hence, $T\alpha = T\beta$ if and only if $T(\alpha - \beta) = 0$. This simplifies enormously the verification that T is 1:1. Let us call a linear transformation T **non-singular** if $T\gamma = 0$ implies

$\gamma = 0$, i.e., if the null space of T is $\{0\}$. Evidently, T is 1:1 if and only if T is non-singular. The extension of this remark is that non-singular linear transformations are those which preserve linear independence.

Theorem 8. *Let* T *be a linear transformation from* V *into* W. *Then* T *is non-singular if and only if* T *carries each linearly independent subset of* V *onto a linearly independent subset of* W.

Proof. First suppose that T is non-singular. Let S be a linearly independent subset of V. If $\alpha_1, \ldots, \alpha_k$ are vectors in S, then the vectors $T\alpha_1, \ldots, T\alpha_k$ are linearly independent; for if

$$c_1(T\alpha_1) + \cdots + c_k(T\alpha_k) = 0$$

then

$$T(c_1\alpha_1 + \cdots + c_k\alpha_k) = 0$$

and since T is non-singular

$$c_1\alpha_1 + \cdots + c_k\alpha_k = 0$$

from which it follows that each $c_i = 0$ because S is an independent set. This argument shows that the image of S under T is independent.

Suppose that T carries independent subsets onto independent subsets. Let α be a non-zero vector in V. Then the set S consisting of the one vector α is independent. The image of S is the set consisting of the one vector $T\alpha$, and this set is independent. Therefore $T\alpha \neq 0$, because the set consisting of the zero vector alone is dependent. This shows that the null space of T is the zero subspace, i.e., T is non-singular. ∎

EXAMPLE 11. Let F be a subfield of the complex numbers (or a field of characteristic zero) and let V be the space of polynomial functions over F. Consider the differentiation operator D and the 'multiplication by x' operator T, from Example 9. Since D sends all constants into 0, D is singular; however, V is not finite dimensional, the range of D is all of V, and it is possible to define a right inverse for D. For example, if E is the indefinite integral operator:

$$E(c_0 + c_1x + \cdots + c_nx^n) = c_0x + \frac{1}{2}c_1x^2 + \cdots + \frac{1}{n+1}c_nx^{n+1}$$

then E is a linear operator on V and $DE = I$. On the other hand, $ED \neq I$ because ED sends the constants into 0. The operator T is in what we might call the reverse situation. If $xf(x) = 0$ for all x, then $f = 0$. Thus T is non-singular and it is possible to find a left inverse for T. For example if U is the operation 'remove the constant term and divide by x':

$$U(c_0 + c_1x + \cdots + c_nx^n) = c_1 + c_2x + \cdots + c_nx^{n-1}$$

then U is a linear operator on V and $UT = I$. But $TU \neq I$ since every

function in the range of TU is in the range of T, which is the space of polynomial functions f such that $f(0) = 0$.

EXAMPLE 12. Let F be a field and let T be the linear operator on F^2 defined by

$$T(x_1, x_2) = (x_1 + x_2, x_1).$$

Then T is non-singular, because if $T(x_1, x_2) = 0$ we have

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_1 &= 0 \end{aligned}$$

so that $x_1 = x_2 = 0$. We also see that T is onto; for, let (z_1, z_2) be any vector in F^2. To show that (z_1, z_2) is in the range of T we must find scalars x_1 and x_2 such that

$$\begin{aligned} x_1 + x_2 &= z_1 \\ x_1 &= z_2 \end{aligned}$$

and the obvious solution is $x_1 = z_2$, $x_2 = z_1 - z_2$. This last computation gives us an explicit formula for T^{-1}, namely,

$$T^{-1}(z_1, z_2) = (z_2, z_1 - z_2).$$

We have seen in Example 11 that a linear transformation may be non-singular without being onto and may be onto without being non-singular. The present example illustrates an important case in which that cannot happen.

Theorem 9. *Let* V *and* W *be finite-dimensional vector spaces over the field* F *such that dim* V = *dim* W. *If* T *is a linear transformation from* V *into* W, *the following are equivalent:*

(i) T *is invertible.*
(ii) T *is non-singular.*
(iii) T *is onto, that is, the range of* T *is* W.

Proof. Let $n = \dim V = \dim W$. From Theorem 2 we know that

$$\text{rank } (T) + \text{nullity } (T) = n.$$

Now T is non-singular if and only if nullity $(T) = 0$, and (since $n = \dim W$) the range of T is W if and only if rank $(T) = n$. Since the rank plus the nullity is n, the nullity is 0 precisely when the rank is n. Therefore T is non-singular if and only if $T(V) = W$. So, if either condition (ii) or (iii) holds, the other is satisfied as well and T is invertible. ∎

We caution the reader not to apply Theorem 9 except in the presence of finite-dimensionality and with $\dim V = \dim W$. Under the hypotheses of Theorem 9, the conditions (i), (ii), and (iii) are also equivalent to these.

(iv) *If* $\{\alpha_1, \ldots, \alpha_n\}$ *is basis for* V, *then* $\{T\alpha_1, \ldots, T\alpha_n\}$ *is a basis for* W.

(v) *There is some basis* $\{\alpha_1, \ldots, \alpha_n\}$ *for* V *such that* $\{T\alpha_1, \ldots, T\alpha_n\}$ *is a basis for* W.

We shall give a proof of the equivalence of the five conditions which contains a different proof that (i), (ii), and (iii) are equivalent.

(i) → (ii). If T is invertible, T is non-singular. (ii) → (iii). Suppose T is non-singular. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis for V. By Theorem 8, $\{T\alpha_1, \ldots, T\alpha_n\}$ is a linearly independent set of vectors in W, and since the dimension of W is also n, this set of vectors is a basis for W. Now let β be any vector in W. There are scalars $c_1, \ldots, c_n$ such that

$$\begin{aligned}\beta &= c_1(T\alpha_1) + \cdots + c_n(T\alpha_n) \\ &= T(c_1\alpha_1 + \cdots + c_n\alpha_n)\end{aligned}$$

which shows that β is in the range of T. (iii) → (iv). We now assume that T is onto. If $\{\alpha_1, \ldots, \alpha_n\}$ is any basis for V, the vectors $T\alpha_1, \ldots, T\alpha_n$ span the range of T, which is all of W by assumption. Since the dimension of W is n, these n vectors must be linearly independent, that is, must comprise a basis for W. (iv) → (v). This requires no comment. (v) → (i). Suppose there is some basis $\{\alpha_1, \ldots, \alpha_n\}$ for V such that $\{T\alpha_1, \ldots, T\alpha_n\}$ is a basis for W. Since the $T\alpha_i$ span W, it is clear that the range of T is all of W. If $\alpha = c_1\alpha_1 + \cdots + c_n\alpha_n$ is in the null space of T, then

$$T(c_1\alpha_1 + \cdots + c_n\alpha_n) = 0$$

or

$$c_1(T\alpha_1) + \cdots + c_n(T\alpha_n) = 0$$

and since the $T\alpha_i$ are independent each $c_i = 0$, and thus $\alpha = 0$. We have shown that the range of T is W, and that T is non-singular, hence T is invertible.

The set of invertible linear operators on a space V, with the operation of composition, provides a nice example of what is known in algebra as a 'group.' Although we shall not have time to discuss groups in any detail, we shall at least give the definition.

Definition. *A* **group** *consists of the following.*

1. *A set* G;

2. *A rule (or operation) which associates with each pair of elements* x, y *in* G *an element* xy *in* G *in such a way that*

   (a) x(yz) = (xy)z, *for all* x, y, *and* z *in* G *(associativity);*

   (b) *there is an element* e *in* G *such that* ex = xe = x, *for every* x *in* G*;*

   (c) *to each element* x *in* G *there corresponds an element* x^{-1} *in* G *such that* $xx^{-1} = x^{-1}x = e$.

We have seen that composition $(U, T) \to UT$ associates with each pair of invertible linear operators on a space V another invertible operator on V. Composition is an associative operation. The identity operator I

satisfies $IT = TI$ for each T, and for an invertible T there is (by Theorem 7) an invertible linear operator T^{-1} such that $TT^{-1} = T^{-1}T = I$. Thus the set of invertible linear operators on V, together with this operation, is a group. The set of invertible $n \times n$ matrices with matrix multiplication as the operation is another example of a group. A group is called **commutative** if it satisfies the condition $xy = yx$ for each x and y. The two examples we gave above are not commutative groups, in general. One often writes the operation in a commutative group as $(x, y) \to x + y$, rather than $(x, y) \to xy$, and then uses the symbol 0 for the 'identity' element e. The set of vectors in a vector space, together with the operation of vector addition, is a commutative group. A field can be described as a set with two operations, called addition and multiplication, which is a commutative group under addition, and in which the non-zero elements form a commutative group under multiplication, with the distributive law $x(y + z) = xy + xz$ holding.

Exercises

1. Let T and U be the linear operators on R^2 defined by

$$T(x_1, x_2) = (x_2, x_1) \quad \text{and} \quad U(x_1, x_2) = (x_1, 0).$$

(a) How would you describe T and U geometrically?

(b) Give rules like the ones defining T and U for each of the transformations $(U + T)$, UT, TU, T^2, U^2.

2. Let T be the (unique) linear operator on C^3 for which

$$T\epsilon_1 = (1, 0, i), \qquad T\epsilon_2 = (0, 1, 1), \qquad T\epsilon_3 = (i, 1, 0).$$

Is T invertible?

3. Let T be the linear operator on R^3 defined by

$$T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3).$$

Is T invertible? If so, find a rule for T^{-1} like the one which defines T.

4. For the linear operator T of Exercise 3, prove that

$$(T^2 - I)(T - 3I) = 0.$$

5. Let $C^{2\times 2}$ be the complex vector space of 2×2 matrices with complex entries. Let

$$B = \begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix}$$

and let T be the linear operator on $C^{2\times 2}$ defined by $T(A) = BA$. What is the rank of T? Can you describe T^2?

6. Let T be a linear transformation from R^3 into R^2, and let U be a linear transformation from R^2 into R^3. Prove that the transformation UT is not invertible. Generalize the theorem.

7. Find two linear operators T and U on R^2 such that $TU = 0$ but $UT \neq 0$.

8. Let V be a vector space over the field F and T a linear operator on V. If $T^2 = 0$, what can you say about the relation of the range of T to the null space of T? Give an example of a linear operator T on R^2 such that $T^2 = 0$ but $T \neq 0$.

9. Let T be a linear operator on the finite-dimensional space V. Suppose there is a linear operator U on V such that $TU = I$. Prove that T is invertible and $U = T^{-1}$. Give an example which shows that this is false when V is not finite-dimensional. (*Hint:* Let $T = D$, the differentiation operator on the space of polynomial functions.)

10. Let A be an $m \times n$ matrix with entries in F and let T be the linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$ defined by $T(X) = AX$. Show that if $m < n$ it may happen that T is onto without being non-singular. Similarly, show that if $m > n$ we may have T non-singular but not onto.

11. Let V be a finite-dimensional vector space and let T be a linear operator on V. Suppose that rank (T^2) = rank (T). Prove that the range and null space of T are disjoint, i.e., have only the zero vector in common.

12. Let p, m, and n be positive integers and F a field. Let V be the space of $m \times n$ matrices over F and W the space of $p \times n$ matrices over F. Let B be a fixed $p \times m$ matrix and let T be the linear transformation from V into W defined by $T(A) = BA$. Prove that T is invertible if and only if $p = m$ and B is an invertible $m \times m$ matrix.

3.3. Isomorphism

If V and W are vector spaces over the field F, any one-one linear transformation T of V onto W is called an **isomorphism of V onto W**. If there exists an isomorphism of V onto W, we say that V is **isomorphic** to W.

Note that V is trivially isomorphic to V, the identity operator being an isomorphism of V onto V. Also, if V is isomorphic to W via an isomorphism T, then W is isomorphic to V, because T^{-1} is an isomorphism of W onto V. The reader should find it easy to verify that if V is isomorphic to W and W is isomorphic to Z, then V is isomorphic to Z. Briefly, isomorphism is an equivalence relation on the class of vector spaces. If there exists an isomorphism of V onto W, we may sometimes say that V and W are isomorphic, rather than V is isomorphic to W. This will cause no confusion because V is isomorphic to W if and only if W is isomorphic to V.

Theorem 10. *Every* n*-dimensional vector space over the field* F *is isomorphic to the space* F^n.

Proof. Let V be an n-dimensional space over the field F and let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ be an ordered basis for V. We define a function T

from V into F^n, as follows: If α is in V, let $T\alpha$ be the n-tuple $(x_1, \ldots, x_n)$ of coordinates of α relative to the ordered basis $\mathcal{B}$, i.e., the n-tuple such that

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n.$$

In our discussion of coordinates in Chapter 2, we verified that this T is linear, one-one, and maps V onto F^n. ∎

For many purposes one often regards isomorphic vector spaces as being 'the same,' although the vectors and operations in the spaces may be quite different, that is, one often identifies isomorphic spaces. We shall not attempt a lengthy discussion of this idea at present but shall let the understanding of isomorphism and the sense in which isomorphic spaces are 'the same' grow as we continue our study of vector spaces.

We shall make a few brief comments. Suppose T is an isomorphism of V onto W. If S is a subset of V, then Theorem 8 tells us that S is linearly independent if and only if the set $T(S)$ in W is independent. Thus in deciding whether S is independent it doesn't matter whether we look at S or $T(S)$. From this one sees that an isomorphism is 'dimension preserving,' that is, any finite-dimensional subspace of V has the same dimension as its image under T. Here is a very simple illustration of this idea. Suppose A is an $m \times n$ matrix over the field F. We have really given two definitions of the solution space of the matrix A. The first is the set of all n-tuples $(x_1, \ldots, x_n)$ in F^n which satisfy each of the equations in the system $AX = 0$. The second is the set of all $n \times 1$ column matrices X such that $AX = 0$. The first solution space is thus a subspace of F^n and the second is a subspace of the space of all $n \times 1$ matrices over F. Now there is a completely obvious isomorphism between F^n and $F^{n\times 1}$, namely,

$$(x_1, \ldots, x_n) \to \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Under this isomorphism, the first solution space of A is carried onto the second solution space. These spaces have the same dimension, and so if we want to prove a theorem about the dimension of the solution space, it is immaterial which space we choose to discuss. In fact, the reader would probably not balk if we chose to identify F^n and the space of $n \times 1$ matrices. We may do this when it is convenient, and when it is not convenient we shall not.

Exercises

1. Let V be the set of complex numbers and let F be the field of real numbers. With the usual operations, V is a vector space over F. Describe explicitly an isomorphism of this space onto R^2.

2. Let V be a vector space over the field of complex numbers, and suppose there is an isomorphism T of V onto C^3. Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be vectors in V such that

$$T\alpha_1 = (1, 0, i), \qquad T\alpha_2 = (-2, 1 + i, 0),$$
$$T\alpha_3 = (-1, 1, 1), \qquad T\alpha_4 = (\sqrt{2}, i, 3).$$

(a) Is α_1 in the subspace spanned by α_2 and α_3?

(b) Let W_1 be the subspace spanned by α_1 and α_2, and let W_2 be the subspace spanned by α_3 and α_4. What is the intersection of W_1 and W_2?

(c) Find a basis for the subspace of V spanned by the four vectors α_j.

3. Let W be the set of all 2×2 complex Hermitian matrices, that is, the set of 2×2 complex matrices A such that $A_{ij} = \overline{A_{ji}}$ (the bar denoting complex conjugation). As we pointed out in Example 6 of Chapter 2, W is a vector space over the field of *real* numbers, under the usual operations. Verify that

$$(x, y, z, t) \to \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

is an isomorphism of R^4 onto W.

4. Show that $F^{m\times n}$ is isomorphic to F^{mn}.

5. Let V be the set of complex numbers regarded as a vector space over the field of real numbers (Exercise 1). We define a function T from V into the space of 2×2 real matrices, as follows. If $z = x + iy$ with x and y real numbers, then

$$T(z) = \begin{bmatrix} x + 7y & 5y \\ -10y & x - 7y \end{bmatrix}.$$

(a) Verify that T is a one-one (real) linear transformation of V into the space of 2×2 real matrices.

(b) Verify that $T(z_1z_2) = T(z_1)T(z_2)$.

(c) How would you describe the range of T?

6. Let V and W be finite-dimensional vector spaces over the field F. Prove that V and W are isomorphic if and only if dim V = dim W.

7. Let V and W be vector spaces over the field F and let U be an isomorphism of V onto W. Prove that $T \to UTU^{-1}$ is an isomorphism of $L(V, V)$ onto $L(W, W)$.

3.4. Representation of Transformations by Matrices

Let V be an n-dimensional vector space over the field F and let W be an m-dimensional vector space over F. Let $\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$ be an ordered basis for V and $\mathfrak{B}' = \{\beta_1, \ldots, \beta_m\}$ an ordered basis for W. If T is any linear transformation from V into W, then T is determined by its action on the vectors α_j. Each of the n vectors $T\alpha_j$ is uniquely expressible as a linear combination

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i \tag{3-3}$$

of the β_i, the scalars $A_{1j}, \ldots, A_{mj}$ being the coordinates of $T\alpha_j$ in the ordered basis $\mathcal{B}'$. Accordingly, the transformation T is determined by the mn scalars A_{ij} via the formulas (3-3). The $m \times n$ matrix A defined by $A(i, j) = A_{ij}$ is called **the matrix of T relative to the pair of ordered bases $\mathcal{B}$ and $\mathcal{B}'$**. Our immediate task is to understand explicitly how the matrix A determines the linear transformation T.

If $\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$ is a vector in V, then

$$\begin{aligned} T\alpha &= T\left(\sum_{j=1}^{n} x_j\alpha_j\right) \\ &= \sum_{j=1}^{n} x_j(T\alpha_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i. \end{aligned}$$

If X is the coordinate matrix of α in the ordered basis $\mathcal{B}$, then the computation above shows that AX is the coordinate matrix of the vector $T\alpha$ in the ordered basis $\mathcal{B}'$, because the scalar

$$\sum_{j=1}^{n} A_{ij}x_j$$

is the entry in the ith row of the column matrix AX. Let us also observe that if A is any $m \times n$ matrix over the field F, then

$$T\left(\sum_{j=1}^{n} x_j\alpha_j\right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i \tag{3-4}$$

defines a linear transformation T from V into W, the matrix of which is A, relative to $\mathcal{B}$, $\mathcal{B}'$. We summarize formally:

Theorem 11. *Let* V *be an* n*-dimensional vector space over the field* F *and* W *an* m*-dimensional vector space over* F. *Let* $\mathcal{B}$ *be an ordered basis for* V *and* $\mathcal{B}'$ *an ordered basis for* W. *For each linear transformation* T *from* V *into* W, *there is an* m $\times$ n *matrix* A *with entries in* F *such that*

$$[\mathrm{T}\alpha]_{\mathcal{B}'} = \mathrm{A}[\alpha]_{\mathcal{B}}$$

for every vector α *in* V. *Furthermore,* $\mathrm{T} \to \mathrm{A}$ *is a one-one correspondence between the set of all linear transformations from* V *into* W *and the set of all* m $\times$ n *matrices over the field* F.

The matrix A which is associated with T in Theorem 11 is called the **matrix of T relative to the ordered bases** $\mathcal{B}$, $\mathcal{B}'$. Note that Equation (3-3) says that A is the matrix whose columns $A_1, \ldots, A_n$ are given by

$$A_j = [T\alpha_j]_{\mathcal{B}'}, \qquad j = 1, \ldots, n.$$

If U is another linear transformation from V into W and $B = [B_1, \ldots, B_n]$ is the matrix of U relative to the ordered bases $\mathcal{B}$, $\mathcal{B}'$ then $cA + B$ is the matrix of $cT + U$ relative to $\mathcal{B}$, $\mathcal{B}'$. That is clear because

$$\begin{aligned} cA_j + B_j &= c[T\alpha_j]_{\mathcal{B}'} + [U\alpha_j]_{\mathcal{B}'} \\ &= [cT\alpha_j + U\alpha_j]_{\mathcal{B}'} \\ &= [(cT + U)\alpha_j]_{\mathcal{B}'}. \end{aligned}$$

Theorem 12. *Let* V *be an* n*-dimensional vector space over the field* F *and let* W *be an* m*-dimensional vector space over* F. *For each pair of ordered bases* $\mathcal{B}$, $\mathcal{B}'$ *for* V *and* W *respectively, the function which assigns to a linear transformation* T *its matrix relative to* $\mathcal{B}$, $\mathcal{B}'$ *is an isomorphism between the space* L(V, W) *and the space of all* m $\times$ n *matrices over the field* F.

Proof. We observed above that the function in question is linear, and as stated in Theorem 11, this function is one-one and maps $L(V, W)$ onto the set of $m \times n$ matrices. ∎

We shall be particularly interested in the representation by matrices of linear transformations of a space into itself, i.e., linear operators on a space V. In this case it is most convenient to use the same ordered basis in each case, that is, to take $\mathcal{B} = \mathcal{B}'$. We shall then call the representing matrix simply the **matrix of T relative to the ordered basis $\mathcal{B}$**. Since this concept will be so important to us, we shall review its definition. If T is a linear operator on the finite-dimensional vector space V and $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ is an ordered basis for V, the matrix of T relative to $\mathcal{B}$ (or, the matrix of T in the ordered basis $\mathcal{B}$) is the $n \times n$ matrix A whose entries A_{ij} are defined by the equations

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i, \qquad j = 1, \ldots, n. \tag{3-5}$$

One must always remember that this matrix representing T depends upon the ordered basis $\mathcal{B}$, and that there is a representing matrix for T in each ordered basis for V. (For transformations of one space into another the matrix depends upon two ordered bases, one for V and one for W.) In order that we shall not forget this dependence, we shall use the notation

$$[T]_{\mathcal{B}}$$

for the matrix of the linear operator T in the ordered basis $\mathcal{B}$. The manner in which this matrix and the ordered basis describe T is that for each α in V

$$[T\alpha]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}.$$

EXAMPLE 13. Let V be the space of $n \times 1$ column matrices over the field F; let W be the space of $m \times 1$ matrices over F; and let A be a fixed $m \times n$ matrix over F. Let T be the linear transformation of V into W defined by $T(X) = AX$. Let $\mathcal{B}$ be the ordered basis for V analogous to the

standard basis in F^n, i.e., the ith vector in $\mathcal{B}$ in the $n \times 1$ matrix X_i with a 1 in row i and all other entries 0. Let $\mathcal{B}'$ be the corresponding ordered basis for W, i.e., the jth vector in $\mathcal{B}'$ is the $m \times 1$ matrix Y_j with a 1 in row j and all other entries 0. Then the matrix of T relative to the pair $\mathcal{B}$, $\mathcal{B}'$ is the matrix A itself. This is clear because the matrix AX_j is the jth column of A.

EXAMPLE 14. Let F be a field and let T be the operator on F^2 defined by

$$T(x_1, x_2) = (x_1, 0).$$

It is easy to see that T is a linear operator on F^2. Let $\mathcal{B}$ be the standard ordered basis for F^2, $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$. Now

$$T\epsilon_1 = T(1, 0) = (1, 0) = 1\epsilon_1 + 0\epsilon_2$$
$$T\epsilon_2 = T(0, 1) = (0, 0) = 0\epsilon_1 + 0\epsilon_2$$

so the matrix of T in the ordered basis $\mathcal{B}$ is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

EXAMPLE 15. Let V be the space of all polynomial functions from R into R of the form

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3$$

that is, the space of polynomial functions of degree three or less. The differentiation operator D of Example 2 maps V into V, since D is 'degree decreasing.' Let $\mathcal{B}$ be the ordered basis for V consisting of the four functions f_1, f_2, f_3, f_4 defined by $f_j(x) = x^{j-1}$. Then

$$\begin{aligned} (Df_1)(x) &= 0, & Df_1 &= 0f_1 + 0f_2 + 0f_3 + 0f_4 \\ (Df_2)(x) &= 1, & Df_2 &= 1f_1 + 0f_2 + 0f_3 + 0f_4 \\ (Df_3)(x) &= 2x, & Df_3 &= 0f_1 + 2f_2 + 0f_3 + 0f_4 \\ (Df_4)(x) &= 3x^2, & Df_4 &= 0f_1 + 0f_2 + 3f_3 + 0f_4 \end{aligned}$$

so that the matrix of D in the ordered basis $\mathcal{B}$ is

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We have seen what happens to representing matrices when transformations are added, namely, that the matrices add. We should now like to ask what happens when we compose transformations. More specifically, let V, W, and Z be vector spaces over the field F of respective dimensions n, m, and p. Let T be a linear transformation from V into W and U a linear transformation from W into Z. Suppose we have ordered bases

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}, \qquad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}, \qquad \mathcal{B}'' = \{\gamma_1, \ldots, \gamma_p\}$$

for the respective spaces V, W, and Z. Let A be the matrix of T relative to the pair $\mathcal{B}$, $\mathcal{B}'$ and let B be the matrix of U relative to the pair $\mathcal{B}'$, $\mathcal{B}''$. It is then easy to see that the matrix C of the transformation UT relative to the pair $\mathcal{B}$, $\mathcal{B}''$ is the product of B and A; for, if α is any vector in V

$$[T\alpha]_{\mathcal{B}'} = A[\alpha]_{\mathcal{B}}$$
$$[U(T\alpha)]_{\mathcal{B}''} = B[T\alpha]_{\mathcal{B}'}$$

and so

$$[(UT)(\alpha)]_{\mathcal{B}''} = BA[\alpha]_{\mathcal{B}}$$

and hence, by the definition and uniqueness of the representing matrix, we must have $C = BA$. One can also see this by carrying out the computation

$$\begin{aligned}(UT)(\alpha_j) &= U(T\alpha_j) \\ &= U\left(\sum_{k=1}^{m} A_{kj}\beta_k\right) \\ &= \sum_{k=1}^{m} A_{kj}(U\beta_k) \\ &= \sum_{k=1}^{m} A_{kj} \sum_{i=1}^{p} B_{ik}\gamma_i \\ &= \sum_{i=1}^{p} \left(\sum_{k=1}^{m} B_{ik}A_{kj}\right)\gamma_i\end{aligned}$$

so that we must have

$$C_{ij} = \sum_{k=1}^{m} B_{ik}A_{kj}. \tag{3-6}$$

We motivated the definition (3-6) of matrix multiplication via operations on the rows of a matrix. One sees here that a very strong motivation for the definition is to be found in composing linear transformations. Let us summarize formally.

Theorem 13. *Let* V, W, *and* Z *be finite-dimensional vector spaces over the field* F; *let* T *be a linear transformation from* V *into* W *and* U *a linear transformation from* W *into* Z. *If* $\mathcal{B}$, $\mathcal{B}'$, *and* $\mathcal{B}''$ *are ordered bases for the spaces* V, W, *and* Z, *respectively, if* A *is the matrix of* T *relative to the pair* $\mathcal{B}$, $\mathcal{B}'$, *and* B *is the matrix of* U *relative to the pair* $\mathcal{B}'$, $\mathcal{B}''$, *then the matrix of the composition* UT *relative to the pair* $\mathcal{B}$, $\mathcal{B}''$ *is the product matrix* C = BA.

We remark that Theorem 13 gives a proof that matrix multiplication is associative—a proof which requires no calculations and is independent of the proof we gave in Chapter 1. We should also point out that we proved a special case of Theorem 13 in Example 12.

It is important to note that if T and U are linear operators on a space V and we are representing by a single ordered basis $\mathcal{B}$, then Theorem 13 assumes the simple form $[UT]_{\mathcal{B}} = [U]_{\mathcal{B}}[T]_{\mathcal{B}}$. Thus in this case, the

correspondence which $\mathcal{B}$ determines between operators and matrices is not only a vector space isomorphism but also preserves products. A simple consequence of this is that the linear operator T is invertible if and only if $[T]_{\mathcal{B}}$ is an invertible matrix. For, the identity operator I is represented by the identity matrix in any ordered basis, and thus

$$UT = TU = I$$

is equivalent to

$$[U]_{\mathcal{B}}[T]_{\mathcal{B}} = [T]_{\mathcal{B}}[U]_{\mathcal{B}} = I.$$

Of course, when T is invertible

$$[T^{-1}]_{\mathcal{B}} = [T]_{\mathcal{B}}^{-1}.$$

Now we should like to inquire what happens to representing matrices when the ordered basis is changed. For the sake of simplicity, we shall consider this question only for linear operators on a space V, so that we can use a single ordered basis. The specific question is this. Let T be a linear operator on the finite-dimensional space V, and let

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{and} \quad \mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$$

be two ordered bases for V. How are the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}$ related? As we observed in Chapter 2, there is a unique (invertible) $n \times n$ matrix P such that

$$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'} \tag{3-7}$$

for every vector α in V. It is the matrix $P = [P_1, \ldots, P_n]$ where $P_j = [\alpha_j']_{\mathcal{B}}$. By definition

$$[T\alpha]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}. \tag{3-8}$$

Applying (3-7) to the vector $T\alpha$, we have

$$[T\alpha]_{\mathcal{B}} = P[T\alpha]_{\mathcal{B}'}. \tag{3-9}$$

Combining (3-7), (3-8), and (3-9), we obtain

$$[T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'} = P[T\alpha]_{\mathcal{B}'}$$

or

$$P^{-1}[T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'} = [T\alpha]_{\mathcal{B}'}$$

and so it must be that

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P. \tag{3-10}$$

This answers our question.

Before stating this result formally, let us observe the following. There is a unique linear operator U which carries $\mathcal{B}$ onto $\mathcal{B}'$, defined by

$$U\alpha_j = \alpha_j', \qquad j = 1, \ldots, n.$$

This operator U is invertible since it carries a basis for V onto a basis for

V. The matrix P (above) is precisely the matrix of the operator U in the ordered basis $\mathfrak{B}$. For, P is defined by

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i$$

and since $U\alpha_j = \alpha'_j$, this equation can be written

$$U\alpha_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

So $P = [U]_{\mathfrak{B}}$, by definition.

Theorem 14. *Let* V *be a finite-dimensional vector space over the field* F, *and let*

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \textit{and} \quad \mathfrak{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

be ordered bases for V. *Suppose* T *is a linear operator on* V. *If* $\mathrm{P} = [\mathrm{P}_1, \ldots, \mathrm{P}_n]$ *is the* n × n *matrix with columns* $\mathrm{P}_j = [\alpha'_j]_{\mathfrak{B}}$, *then*

$$[\mathrm{T}]_{\mathfrak{B}'} = \mathrm{P}^{-1}[\mathrm{T}]_{\mathfrak{B}}\mathrm{P}.$$

Alternatively, if U *is the invertible operator on* V *defined by* $\mathrm{U}\alpha_j = \alpha'_j$, $j = 1, \ldots, n$, *then*

$$[\mathrm{T}]_{\mathfrak{B}'} = [\mathrm{U}]_{\mathfrak{B}}^{-1}[\mathrm{T}]_{\mathfrak{B}}[\mathrm{U}]_{\mathfrak{B}}.$$

EXAMPLE 16. Let T be the linear operator on R^2 defined by $T(x_1, x_2) = (x_1, 0)$. In Example 14 we showed that the matrix of T in the standard ordered basis $\mathfrak{B} = \{\epsilon_1, \epsilon_2\}$ is

$$[T]_{\mathfrak{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Suppose $\mathfrak{B}'$ is the ordered basis for R^2 consisting of the vectors $\epsilon'_1 = (1, 1)$, $\epsilon'_2 = (2, 1)$. Then

$$\begin{aligned} \epsilon'_1 &= \epsilon_1 + \epsilon_2 \\ \epsilon'_2 &= 2\epsilon_1 + \epsilon_2 \end{aligned}$$

so that P is the matrix

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}.$$

By a short computation

$$P^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}.$$

Thus

$$\begin{aligned} [T]_{\mathfrak{B}'} &= P^{-1}[T]_{\mathfrak{B}}P \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}. \end{aligned}$$

We can easily check that this is correct because

$$T\epsilon_1' = (1, 0) = -\epsilon_1' + \epsilon_2'$$
$$T\epsilon_2' = (2, 0) = -2\epsilon_1' + 2\epsilon_2'.$$

EXAMPLE 17. Let V be the space of polynomial functions from R into R which have 'degree' less than or equal to 3. As in Example 15, let D be the differentiation operator on V, and let

$$\mathcal{B} = \{f_1, f_2, f_3, f_4\}$$

be the ordered basis for V defined by $f_i(x) = x^{i-1}$. Let t be a real number and define $g_i(x) = (x + t)^{i-1}$, that is

$$g_1 = f_1$$
$$g_2 = tf_1 + f_2$$
$$g_3 = t^2f_1 + 2tf_2 + f_3$$
$$g_4 = t^3f_1 + 3t^2f_2 + 3tf_3 + f_4.$$

Since the matrix

$$P = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is easily seen to be invertible with

$$P^{-1} = \begin{bmatrix} 1 & -t & t^2 & -t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

it follows that $\mathcal{B}' = \{g_1, g_2, g_3, g_4\}$ is an ordered basis for V. In Example 15, we found that the matrix of D in the ordered basis $\mathcal{B}$ is

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrix of D in the ordered basis $\mathcal{B}'$ is thus

$$P^{-1}[D]_{\mathcal{B}}P = \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 2 & 6t \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus D is represented by the same matrix in the ordered bases $\mathfrak{B}$ and $\mathfrak{B}'$. Of course, one can see this somewhat more directly since

$$\begin{aligned} Dg_1 &= 0 \\ Dg_2 &= g_1 \\ Dg_3 &= 2g_2 \\ Dg_4 &= 3g_3. \end{aligned}$$

This example illustrates a good point. If one knows the matrix of a linear operator in some ordered basis $\mathfrak{B}$ and wishes to find the matrix in another ordered basis $\mathfrak{B}'$, it is often most convenient to perform the coordinate change using the invertible matrix P; however, it may be a much simpler task to find the representing matrix by a direct appeal to its definition.

Definition. *Let* A *and* B *be* n × n *(square) matrices over the field* F. *We say that* B **is similar to** A **over** F *if there is an invertible* n × n *matrix* P *over* F *such that* $B = P^{-1}AP$.

According to Theorem 14, we have the following: If V is an n-dimensional vector space over F and $\mathfrak{B}$ and $\mathfrak{B}'$ are two ordered bases for V, then for each linear operator T on V the matrix $B = [T]_{\mathfrak{B}'}$ is similar to the matrix $A = [T]_{\mathfrak{B}}$. The argument also goes in the other direction. Suppose A and B are $n \times n$ matrices and that B is similar to A. Let V be any n-dimensional space over F and let $\mathfrak{B}$ be an ordered basis for V. Let T be the linear operator on V which is represented in the basis $\mathfrak{B}$ by A. If $B = P^{-1}AP$, let $\mathfrak{B}'$ be the ordered basis for V obtained from $\mathfrak{B}$ by P, i.e.,

$$\alpha_j' = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Then the matrix of T in the ordered basis $\mathfrak{B}'$ will be B.

Thus the statement that B is similar to A means that on each n-dimensional space over F the matrices A and B represent the same linear transformation in two (possibly) different ordered bases.

Note that each $n \times n$ matrix A is similar to itself, using $P = I$; if B is similar to A, then A is similar to B, for $B = P^{-1}AP$ implies that $A = (P^{-1})^{-1}BP^{-1}$; if B is similar to A and C is similar to B, then C is similar to A, for $B = P^{-1}AP$ and $C = Q^{-1}BQ$ imply that $C = (PQ)^{-1}A(PQ)$. Thus, similarity is an equivalence relation on the set of $n \times n$ matrices over the field F. Also note that the only matrix similar to the identity matrix I is I itself, and that the only matrix similar to the zero matrix is the zero matrix itself.

Exercises

1. Let T be the linear operator on C^2 defined by $T(x_1, x_2) = (x_1, 0)$. Let $\mathcal{B}$ be the standard ordered basis for C^2 and let $\mathcal{B}' = \{\alpha_1, \alpha_2\}$ be the ordered basis defined by $\alpha_1 = (1, i)$, $\alpha_2 = (-i, 2)$.

(a) What is the matrix of T relative to the pair $\mathcal{B}$, $\mathcal{B}'$?

(b) What is the matrix of T relative to the pair $\mathcal{B}'$, $\mathcal{B}$?

(c) What is the matrix of T in the ordered basis $\mathcal{B}'$?

(d) What is the matrix of T in the ordered basis $\{\alpha_2, \alpha_1\}$?

2. Let T be the linear transformation from R^3 into R^2 defined by

$$T(x_1, x_2, x_3) = (x_1 + x_2, 2x_3 - x_1).$$

(a) If $\mathcal{B}$ is the standard ordered basis for R^3 and $\mathcal{B}'$ is the standard ordered basis for R^2, what is the matrix of T relative to the pair $\mathcal{B}$, $\mathcal{B}'$?

(b) If $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ and $\mathcal{B}' = \{\beta_1, \beta_2\}$, where

$$\alpha_1 = (1, 0, -1), \quad \alpha_2 = (1, 1, 1), \quad \alpha_3 = (1, 0, 0), \quad \beta_1 = (0, 1), \quad \beta_2 = (1, 0)$$

what is the matrix of T relative to the pair $\mathcal{B}$, $\mathcal{B}'$?

3. Let T be a linear operator on F^n, let A be the matrix of T in the standard ordered basis for F^n, and let W be the subspace of F^n spanned by the column vectors of A. What does W have to do with T?

4. Let V be a two-dimensional vector space over the field F, and let $\mathcal{B}$ be an ordered basis for V. If T is a linear operator on V and

$$[T]_{\mathcal{B}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

prove that $T^2 - (a + d)T + (ad - bc)I = 0$.

5. Let T be the linear operator on R^3, the matrix of which in the standard ordered basis is

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 4 \end{bmatrix}.$$

Find a basis for the range of T and a basis for the null space of T.

6. Let T be the linear operator on R^2 defined by

$$T(x_1, x_2) = (-x_2, x_1).$$

(a) What is the matrix of T in the standard ordered basis for R^2?

(b) What is the matrix of T in the ordered basis $\mathcal{B} = \{\alpha_1, \alpha_2\}$, where $\alpha_1 = (1, 2)$ and $\alpha_2 = (1, -1)$?

(c) Prove that for every real number c the operator $(T - cI)$ is invertible.

(d) Prove that if $\mathcal{B}$ is any ordered basis for R^2 and $[T]_{\mathcal{B}} = A$, then $A_{12}A_{21} \neq 0$.

7. Let T be the linear operator on R^3 defined by

$$T(x_1, x_2, x_3) = (3x_1 + x_3, -2x_1 + x_2, -x_1 + 2x_2 + 4x_3).$$

(a) What is the matrix of T in the standard ordered basis for R^3?

(b) What is the matrix of T in the ordered basis

$$\{\alpha_1, \alpha_2, \alpha_3\}$$

where $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (-1, 2, 1)$, and $\alpha_3 = (2, 1, 1)$?

(c) Prove that T is invertible and give a rule for T^{-1} like the one which defines T.

8. Let θ be a real number. Prove that the following two matrices are similar over the field of complex numbers:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$

(*Hint:* Let T be the linear operator on C^2 which is represented by the first matrix in the standard ordered basis. Then find vectors α_1 and α_2 such that $T\alpha_1 = e^{i\theta}\alpha_1$, $T\alpha_2 = e^{-i\theta}\alpha_2$, and $\{\alpha_1, \alpha_2\}$ is a basis.)

9. Let V be a finite-dimensional vector space over the field F and let S and T be linear operators on V. We ask: When do there exist ordered bases $\mathcal{B}$ and $\mathcal{B}'$ for V such that $[S]_{\mathcal{B}} = [T]_{\mathcal{B}'}$? Prove that such bases exist if and only if there is an invertible linear operator U on V such that $T = USU^{-1}$. (*Outline of proof:* If $[S]_{\mathcal{B}} = [T]_{\mathcal{B}'}$, let U be the operator which carries $\mathcal{B}$ onto $\mathcal{B}'$ and show that $S = UTU^{-1}$. Conversely, if $T = USU^{-1}$ for some invertible U, let $\mathcal{B}$ be any ordered basis for V and let $\mathcal{B}'$ be its image under U. Then show that $[S]_{\mathcal{B}} = [T]_{\mathcal{B}'}$.)

10. We have seen that the linear operator T on R^2 defined by $T(x_1, x_2) = (x_1, 0)$ is represented in the standard ordered basis by the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

This operator satisfies $T^2 = T$. Prove that if S is a linear operator on R^2 such that $S^2 = S$, then $S = 0$, or $S = I$, or there is an ordered basis $\mathcal{B}$ for R^2 such that $[S]_{\mathcal{B}} = A$ (above).

11. Let W be the space of all $n \times 1$ column matrices over a field F. If A is an $n \times n$ matrix over F, then A defines a linear operator L_A on W through left multiplication: $L_A(X) = AX$. Prove that every linear operator on W is left multiplication by some $n \times n$ matrix, i.e., is L_A for some A.

Now suppose V is an n-dimensional vector space over the field F, and let $\mathcal{B}$ be an ordered basis for V. For each α in V, define $U\alpha = [\alpha]_{\mathcal{B}}$. Prove that U is an isomorphism of V onto W. If T is a linear operator on V, then UTU^{-1} is a linear operator on W. Accordingly, UTU^{-1} is left multiplication by some $n \times n$ matrix A. What is A?

12. Let V be an n-dimensional vector space over the field F, and let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ be an ordered basis for V.

(a) According to Theorem 1, there is a unique linear operator T on V such that

$$T\alpha_j = \alpha_{j+1}, \qquad j = 1, \ldots, n-1, \qquad T\alpha_n = 0.$$

What is the matrix A of T in the ordered basis $\mathcal{B}$?

(b) Prove that $T^n = 0$ but $T^{n-1} \neq 0$.

(c) Let S be any linear operator on V such that $S^n = 0$ but $S^{n-1} \neq 0$. Prove that there is an ordered basis $\mathcal{B}'$ for V such that the matrix of S in the ordered basis $\mathcal{B}'$ is the matrix A of part (a).

(d) Prove that if M and N are $n \times n$ matrices over F such that $M^n = N^n = 0$ but $M^{n-1} \neq 0 \neq N^{n-1}$, then M and N are similar.

13. Let V and W be finite-dimensional vector spaces over the field F and let T be a linear transformation from V into W. If

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{and} \quad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$$

are ordered bases for V and W, respectively, define the linear transformations $E^{p,q}$ as in the proof of Theorem 5: $E^{p,q}(\alpha_i) = \delta_{iq}\beta_p$. Then the $E^{p,q}$, $1 \leq p \leq m$, $1 \leq q \leq n$, form a basis for $L(V, W)$, and so

$$T = \sum_{p=1}^{m} \sum_{q=1}^{n} A_{pq} E^{p,q}$$

for certain scalars A_{pq} (the coordinates of T in this basis for $L(V, W)$). Show that the matrix A with entries $A(p, q) = A_{pq}$ is precisely the matrix of T relative to the pair $\mathcal{B}$, $\mathcal{B}'$.

3.5. Linear Functionals

If V is a vector space over the field F, a linear transformation f from V into the scalar field F is also called a **linear functional** on V. If we start from scratch, this means that f is a function from V into F such that

$$f(c\alpha + \beta) = cf(\alpha) + f(\beta)$$

for all vectors α and β in V and all scalars c in F. The concept of linear functional is important in the study of finite-dimensional spaces because it helps to organize and clarify the discussion of subspaces, linear equations, and coordinates.

EXAMPLE 18. Let F be a field and let $a_1, \ldots, a_n$ be scalars in F. Define a function f on F^n by

$$f(x_1, \ldots, x_n) = a_1x_1 + \cdots + a_nx_n.$$

Then f is a linear functional on F^n. It is the linear functional which is represented by the matrix $[a_1 \cdots a_n]$ relative to the standard ordered basis for F^n and the basis $\{1\}$ for F:

$$a_j = f(\epsilon_j), \qquad j = 1, \ldots, n.$$

Every linear functional on F^n is of this form, for some scalars $a_1, \ldots, a_n$. That is immediate from the definition of linear functional because we define $a_j = f(\epsilon_j)$ and use the linearity

$$\begin{aligned} f(x_1, \ldots, x_n) &= f\left(\sum_j x_j\epsilon_j\right) \\ &= \sum_j x_j f(\epsilon_j) \\ &= \sum_j a_j x_j. \end{aligned}$$

EXAMPLE 19. Here is an important example of a linear functional. Let n be a positive integer and F a field. If A is an $n \times n$ matrix with entries in F, the **trace** of A is the scalar

$$\operatorname{tr} A = A_{11} + A_{22} + \cdots + A_{nn}.$$

The trace function is a linear functional on the matrix space $F^{n\times n}$ because

$$\begin{aligned}\operatorname{tr}(cA + B) &= \sum_{i=1}^{n} (cA_{ii} + B_{ii}) \\ &= c \sum_{i=1}^{n} A_{ii} + \sum_{i=1}^{n} B_{ii} \\ &= c \operatorname{tr} A + \operatorname{tr} B.\end{aligned}$$

EXAMPLE 20. Let V be the space of all polynomial functions from the field F into itself. Let t be an element of F. If we define

$$L_t(p) = p(t)$$

then L_t is a linear functional on V. One usually describes this by saying that, for each t, 'evaluation at t' is a linear functional on the space of polynomial functions. Perhaps we should remark that the fact that the functions are polynomials plays no role in this example. Evaluation at t is a linear functional on the space of all functions from F into F.

EXAMPLE 21. This may be the most important linear functional in mathematics. Let $[a, b]$ be a closed interval on the real line and let $C([a, b])$ be the space of continuous real-valued functions on $[a, b]$. Then

$$L(g) = \int_a^b g(t)\, dt$$

defines a linear functional L on $C([a, b])$.

If V is a vector space, the collection of all linear functionals on V forms a vector space in a natural way. It is the space $L(V, F)$. We denote this space by V^* and call it the **dual space** of V:

$$V^* = L(V, F).$$

If V is finite-dimensional, we can obtain a rather explicit description of the dual space V^*. From Theorem 5 we know something about the space V^*, namely that

$$\dim V^* = \dim V.$$

Let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ be a basis for V. According to Theorem 1, there is (for each i) a unique linear functional f_i on V such that

$$f_i(\alpha_j) = \delta_{ij}. \tag{3-11}$$

In this way we obtain from $\mathcal{B}$ a set of n distinct linear functionals $f_1, \ldots, f_n$ on V. These functionals are also linearly independent. For, suppose

(3-12) $$f = \sum_{i=1}^{n} c_i f_i.$$

Then

$$f(\alpha_j) = \sum_{i=1}^{n} c_i f_i(\alpha_j)$$

$$= \sum_{i=1}^{n} c_i \delta_{ij}$$

$$= c_j.$$

In particular, if f is the zero functional, $f(\alpha_j) = 0$ for each j and hence the scalars c_j are all 0. Now $f_1, \ldots, f_n$ are n linearly independent functionals, and since we know that V^* has dimension n, it must be that $\mathcal{B}^* = \{f_1, \ldots, f_n\}$ is a basis for V^*. This basis is called the **dual basis** of $\mathcal{B}$.

Theorem 15. *Let* V *be a finite-dimensional vector space over the field* F, *and let* $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ *be a basis for* V. *Then there is a unique dual basis* $\mathcal{B}^* = \{f_1, \ldots, f_n\}$ *for* V^* *such that* $f_i(\alpha_j) = \delta_{ij}$. *For each linear functional* f *on* V *we have*

(3-13) $$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

and for each vector α *in* V *we have*

(3-14) $$\alpha = \sum_{i=1}^{n} f_i(\alpha) \alpha_i.$$

Proof. We have shown above that there is a unique basis which is 'dual' to $\mathcal{B}$. If f is a linear functional on V, then f is some linear combination (3-12) of the f_i, and as we observed after (3-12) the scalars c_j must be given by $c_j = f(\alpha_j)$. Similarly, if

$$\alpha = \sum_{i=1}^{n} x_i \alpha_i$$

is a vector in V, then

$$f_j(\alpha) = \sum_{i=1}^{n} x_i f_j(\alpha_i)$$

$$= \sum_{i=1}^{n} x_i \delta_{ij}$$

$$= x_j$$

so that the unique expression for α as a linear combination of the α_i is

$$\alpha = \sum_{i=1}^{n} f_i(\alpha) \alpha_i. \quad \blacksquare$$

Equation (3-14) provides us with a nice way of describing what the dual basis is. It says, if $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ is an ordered basis for V and

$\mathcal{B}^* = \{f_1, \ldots, f_n\}$ is the dual basis, then f_i is precisely the function which assigns to each vector α in V the ith coordinate of α relative to the ordered basis $\mathcal{B}$. Thus we may also call the f_i the coordinate functions for $\mathcal{B}$. The formula (3-13), when combined with (3-14) tells us the following: If f is in V^*, and we let $f(\alpha_i) = a_i$, then when

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$$

we have

$$f(\alpha) = a_1x_1 + \cdots + a_nx_n. \tag{3-15}$$

In other words, if we choose an ordered basis $\mathcal{B}$ for V and describe each vector in V by its n-tuple of coordinates $(x_1, \ldots, x_n)$ relative to $\mathcal{B}$, then every linear functional on V has the form (3-15). This is the natural generalization of Example 18, which is the special case $V = F^n$ and $\mathcal{B} = \{\epsilon_1, \ldots, \epsilon_n\}$.

EXAMPLE 22. Let V be the vector space of all polynomial functions from R into R which have degree less than or equal to 2. Let t_1, t_2, and t_3 be any three *distinct* real numbers, and let

$$L_i(p) = p(t_i).$$

Then L_1, L_2, and L_3 are linear functionals on V. These functionals are linearly independent; for, suppose

$$L = c_1L_1 + c_2L_2 + c_3L_3.$$

If $L = 0$, i.e., if $L(p) = 0$ for each p in V, then applying L to the particular polynomial 'functions' 1, x, x^2, we obtain

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ t_1c_1 + t_2c_2 + t_3c_3 &= 0 \\ t_1^2c_1 + t_2^2c_2 + t_3^2c_3 &= 0 \end{aligned}$$

From this it follows that $c_1 = c_2 = c_3 = 0$, because (as a short computation shows) the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \\ t_1^2 & t_2^2 & t_3^2 \end{bmatrix}$$

is invertible when t_1, t_2, and t_3 are distinct. Now the L_i are independent, and since V has dimension 3, these functionals form a basis for V^*. What is the basis for V, of which this is the dual? Such a basis $\{p_1, p_2, p_3\}$ for V must satisfy

$$L_i(p_j) = \delta_{ij}$$

or

$$p_j(t_i) = \delta_{ij}.$$

These polynomial functions are rather easily seen to be

$$p_1(x) = \frac{(x - t_2)(x - t_3)}{(t_1 - t_2)(t_1 - t_3)}$$

$$p_2(x) = \frac{(x - t_1)(x - t_3)}{(t_2 - t_1)(t_2 - t_3)}$$

$$p_3(x) = \frac{(x - t_1)(x - t_2)}{(t_3 - t_1)(t_3 - t_2)}.$$

The basis $\{p_1, p_2, p_3\}$ for V is interesting, because according to (3-14) we have for each p in V

$$p = p(t_1)p_1 + p(t_2)p_2 + p(t_3)p_3.$$

Thus, if c_1, c_2, and c_3 are any real numbers, there is exactly one polynomial function p over R which has degree at most 2 and satisfies $p(t_j) = c_j$, $j = 1, 2, 3$. This polynomial function is $p = c_1p_1 + c_2p_2 + c_3p_3$.

Now let us discuss the relationship between linear functionals and subspaces. If f is a non-zero linear functional, then the rank of f is 1 because the range of f is a non-zero subspace of the scalar field and must (therefore) be the scalar field. If the underlying space V is finite-dimensional, the rank plus nullity theorem (Theorem 2) tells us that the null space N_f has dimension

$$\dim N_f = \dim V - 1.$$

In a vector space of dimension n, a subspace of dimension $n - 1$ is called a **hyperspace.** Such spaces are sometimes called hyperplanes or subspaces of codimension 1. Is every hyperspace the null space of a linear functional? The answer is easily seen to be yes. It is not much more difficult to show that each d-dimensional subspace of an n-dimensional space is the intersection of the null spaces of $(n - d)$ linear functionals (Theorem 16 below).

Definition. *If* V *is a vector space over the field* F *and* S *is a subset of* V, *the* **annihilator** *of* S *is the set* S^0 *of linear functionals* f *on* V *such that* $f(\alpha) = 0$ *for every* α *in* S.

It should be clear to the reader that S^0 is a subspace of V^*, whether S is a subspace of V or not. If S is the set consisting of the zero vector alone, then $S^0 = V^*$. If $S = V$, then S^0 is the zero subspace of V^*. (This is easy to see when V is finite-dimensional.)

Theorem 16. *Let* V *be a finite-dimensional vector space over the field* F, *and let* W *be a subspace of* V. *Then*

$$dim\ W + dim\ W^0 = dim\ V.$$

Proof. Let k be the dimension of W and $\{\alpha_1, \ldots, \alpha_k\}$ a basis for W. Choose vectors $\alpha_{k+1}, \ldots, \alpha_n$ in V such that $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V. Let $\{f_1, \ldots, f_n\}$ be the basis for V^* which is dual to this basis for V.

The claim is that $\{f_{k+1}, \ldots, f_n\}$ is a basis for the annihilator W^0. Certainly f_i belongs to W^0 for $i \geq k+1$, because

$$f_i(\alpha_j) = \delta_{ij}$$

and $\delta_{ij} = 0$ if $i \geq k+1$ and $j \leq k$; from this it follows that, for $i \geq k+1$, $f_i(\alpha) = 0$ whenever α is a linear combination of $\alpha_1, \ldots, \alpha_k$. The functionals $f_{k+1}, \ldots, f_n$ are independent, so all we must show is that they span W^0. Suppose f is in V^*. Now

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

so that if f is in W^0 we have $f(\alpha_i) = 0$ for $i \leq k$ and

$$f = \sum_{i=k+1}^{n} f(\alpha_i) f_i.$$

We have shown that if dim $W = k$ and dim $V = n$ then dim $W^0 = n - k$. ∎

Corollary. *If* W *is a* k*-dimensional subspace of an* n*-dimensional vector space* V, *then* W *is the intersection of* (n − k) *hyperspaces in* V.

Proof. This is a corollary of the proof of Theorem 16 rather than its statement. In the notation of the proof, W is exactly the set of vectors α such that $f_i(\alpha) = 0$, $i = k+1, \ldots, n$. In case $k = n-1$, W is the null space of f_n. ∎

Corollary. *If* W_1 *and* W_2 *are subspaces of a finite-dimensional vector space, then* $W_1 = W_2$ *if and only if* $W_1^0 = W_2^0$.

Proof. If $W_1 = W_2$, then of course $W_1^0 = W_2^0$. If $W_1 \neq W_2$, then one of the two subspaces contains a vector which is not in the other. Suppose there is a vector α which is in W_2 but not in W_1. By the previous corollaries (or the proof of Theorem 16) there is a linear functional f such that $f(\beta) = 0$ for all β in W, but $f(\alpha) \neq 0$. Then f is in W_1^0 but not in W_2^0 and $W_1^0 \neq W_2^0$. ∎

In the next section we shall give different proofs for these two corollaries. The first corollary says that, if we select some ordered basis for the space, each k-dimensional subspace can be described by specifying $(n-k)$ homogeneous linear conditions on the coordinates relative to that basis.

Let us look briefly at systems of homogeneous linear equations from the point of view of linear functionals. Suppose we have a system of linear equations,

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= 0 \\ &\vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= 0 \end{aligned}$$

for which we wish to find the solutions. If we let f_i, $i = 1, \ldots, m$, be the linear functional on F^n defined by

$$f_i(x_1, \ldots, x_n) = A_{i1}x_1 + \cdots + A_{in}x_n$$

then we are seeking the subspace of F^n of all α such that

$$f_i(\alpha) = 0, \qquad i = 1, \ldots, m.$$

In other words, we are seeking the subspace annihilated by $f_1, \ldots, f_m$. Row-reduction of the coefficient matrix provides us with a systematic method of finding this subspace. The n-tuple $(A_{i1}, \ldots, A_{in})$ gives the coordinates of the linear functional f_i relative to the basis which is dual to the standard basis for F^n. The row space of the coefficient matrix may thus be regarded as the space of linear functionals spanned by $f_1, \ldots, f_m$. The solution space is the subspace annihilated by this space of functionals.

Now one may look at the system of equations from the 'dual' point of view. That is, suppose that we are given m vectors in F^n

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

and we wish to find the annihilator of the subspace spanned by these vectors. Since a typical linear functional on F^n has the form

$$f(x_1, \ldots, x_n) = c_1x_1 + \cdots + c_nx_n$$

the condition that f be in this annihilator is that

$$\sum_{j=1}^{n} A_{ij}c_j = 0, \qquad i = 1, \ldots, m$$

that is, that $(c_1, \ldots, c_n)$ be a solution of the system $AX = 0$. From this point of view, row-reduction gives us a systematic method of finding the annihilator of the subspace spanned by a given finite set of vectors in F^n.

EXAMPLE 23. Here are three linear functionals on R^4:

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_2 + 2x_3 + x_4 \\ f_2(x_1, x_2, x_3, x_4) &= 2x_2 + x_4 \\ f_3(x_1, x_2, x_3, x_4) &= -2x_1 - 4x_3 + 3x_4. \end{aligned}$$

The subspace which they annihilate may be found explicitly by finding the row-reduced echelon form of the matrix

$$A = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix}.$$

A short calculation, or a peek at Example 21 of Chapter 2, shows that

$$R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Therefore, the linear functionals

$$\begin{aligned} g_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_3 \\ g_2(x_1, x_2, x_3, x_4) &= x_2 \\ g_3(x_1, x_2, x_3, x_4) &= x_4 \end{aligned}$$

span the same subspace of $(R^4)^*$ and annihilate the same subspace of R^4 as do f_1, f_2, f_3. The subspace annihilated consists of the vectors with

$$\begin{aligned} x_1 &= -2x_3 \\ x_2 &= x_4 = 0. \end{aligned}$$

EXAMPLE 24. Let W be the subspace of R^5 which is spanned by the vectors

$$\begin{aligned} \alpha_1 &= (2, -2, 3, 4, -1), & \alpha_3 &= (0, 0, -1, -2, 3) \\ \alpha_2 &= (-1, 1, 2, 5, 2), & \alpha_4 &= (1, -1, 2, 3, 0). \end{aligned}$$

How does one describe W^0, the annihilator of W? Let us form the 4×5 matrix A with row vectors $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, and find the row-reduced echelon matrix R which is row-equivalent to A:

$$A = \begin{bmatrix} 2 & -2 & 3 & 4 & -1 \\ -1 & 1 & 2 & 5 & 2 \\ 0 & 0 & -1 & -2 & 3 \\ 1 & -1 & 2 & 3 & 0 \end{bmatrix} \longrightarrow R = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

If f is a linear functional on R^5:

$$f(x_1, \ldots, x_5) = \sum_{j=1}^{5} c_j x_j$$

then f is in W^0 if and only if $f(\alpha_i) = 0$, $i = 1, 2, 3, 4$, i.e., if and only if

$$\sum_{j=1}^{5} A_{ij} c_j = 0, \qquad 1 \leq i \leq 4.$$

This is equivalent to

$$\sum_{j=1}^{5} R_{ij} c_j = 0, \qquad 1 \leq i \leq 3$$

or

$$\begin{aligned} c_1 - c_2 - c_4 &= 0 \\ c_3 + 2c_4 &= 0 \\ c_5 &= 0. \end{aligned}$$

We obtain all such linear functionals f by assigning arbitrary values to c_2 and c_4, say $c_2 = a$ and $c_4 = b$, and then finding the corresponding $c_1 = a + b$, $c_3 = -2b$, $c_5 = 0$. So W^0 consists of all linear functionals f of the form

$$f(x_1, x_2, x_3, x_4, x_5) = (a + b)x_1 + ax_2 - 2bx_3 + bx_4.$$

The dimension of W^0 is 2 and a basis $\{f_1, f_2\}$ for W^0 can be found by first taking $a = 1$, $b = 0$ and then $a = 0$, $b = 1$:

$$f_1(x_1, \ldots, x_5) = x_1 + x_2$$
$$f_2(x_1, \ldots, x_5) = x_1 - 2x_3 + x_4.$$

The above general f in W^0 is $f = af_1 + bf_2$.

Exercises

1. In R^3, let $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (0, 1, -2)$, $\alpha_3 = (-1, -1, 0)$.

(a) If f is a linear functional on R^3 such that

$$f(\alpha_1) = 1, \qquad f(\alpha_2) = -1, \qquad f(\alpha_3) = 3,$$

and if $\alpha = (a, b, c)$, find $f(\alpha)$.

(b) Describe explicitly a linear functional f on R^3 such that

$$f(\alpha_1) = f(\alpha_2) = 0 \quad \text{but} \quad f(\alpha_3) \neq 0.$$

(c) Let f be any linear functional such that

$$f(\alpha_1) = f(\alpha_2) = 0 \quad \text{and} \quad f(\alpha_3) \neq 0.$$

If $\alpha = (2, 3, -1)$, show that $f(\alpha) \neq 0$.

2. Let $\mathfrak{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ be the basis for C^3 defined by

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 1, 1), \qquad \alpha_3 = (2, 2, 0).$$

Find the dual basis of $\mathfrak{B}$.

3. If A and B are $n \times n$ matrices over the field F, show that trace (AB) = trace (BA). Now show that similar matrices have the same trace.

4. Let V be the vector space of all polynomial functions p from R into R which have degree 2 or less:

$$p(x) = c_0 + c_1 x + c_2 x^2.$$

Define three linear functionals on V by

$$f_1(p) = \int_0^1 p(x)\, dx, \qquad f_2(p) = \int_0^2 p(x)\, dx, \qquad f_3(p) = \int_0^{-1} p(x)\, dx.$$

Show that $\{f_1, f_2, f_3\}$ is a basis for V^* by exhibiting the basis for V of which it is the dual.

5. If A and B are $n \times n$ complex matrices, show that $AB - BA = I$ is impossible.

6. Let m and n be positive integers and F a field. Let $f_1, \ldots, f_m$ be linear functionals on F^n. For α in F^n define

$$T\alpha = (f_1(\alpha), \ldots, f_m(\alpha)).$$

Show that T is a linear transformation from F^n into F^m. Then show that every linear transformation from F^n into F^m is of the above form, for some $f_1, \ldots, f_m$.

7. Let $\alpha_1 = (1, 0, -1, 2)$ and $\alpha_2 = (2, 3, 1, 1)$, and let W be the subspace of R^4 spanned by α_1 and α_2. Which linear functionals f:

$$f(x_1, x_2, x_3, x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$$

are in the annihilator of W?

8. Let W be the subspace of R^5 which is spanned by the vectors

$$\alpha_1 = \epsilon_1 + 2\epsilon_2 + \epsilon_3, \qquad \alpha_2 = \epsilon_2 + 3\epsilon_3 + 3\epsilon_4 + \epsilon_5$$
$$\alpha_3 = \epsilon_1 + 4\epsilon_2 + 6\epsilon_3 + 4\epsilon_4 + \epsilon_5.$$

Find a basis for W^0.

9. Let V be the vector space of all 2×2 matrices over the field of real numbers, and let

$$B = \begin{bmatrix} 2 & -2 \\ -1 & 1 \end{bmatrix}.$$

Let W be the subspace of V consisting of all A such that $AB = 0$. Let f be a linear functional on V which is in the annihilator of W. Suppose that $f(I) = 0$ and $f(C) = 3$, where I is the 2×2 identity matrix and

$$C = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Find $f(B)$.

10. Let F be a subfield of the complex numbers. We define n linear functionals on F^n $(n \geq 2)$ by

$$f_k(x_1, \ldots, x_n) = \sum_{j=1}^{n} (k - j)x_j, \qquad 1 \leq k \leq n.$$

What is the dimension of the subspace annihilated by $f_1, \ldots, f_n$?

11. Let W_1 and W_2 be subspaces of a finite-dimensional vector space V.
(a) Prove that $(W_1 + W_2)^0 = W_1^0 \cap W_2^0$.
(b) Prove that $(W_1 \cap W_2)^0 = W_1^0 + W_2^0$.

12. Let V be a finite-dimensional vector space over the field F and let W be a subspace of V. If f is a linear functional on W, prove that there is a linear functional g on V such that $g(\alpha) = f(\alpha)$ for each α in the subspace W.

13. Let F be a subfield of the field of complex numbers and let V be any vector space over F. Suppose that f and g are linear functionals on V such that the function h defined by $h(\alpha) = f(\alpha)g(\alpha)$ is also a linear functional on V. Prove that either $f = 0$ or $g = 0$.

14. Let F be a field of characteristic zero and let V be a finite-dimensional vector space over F. If $\alpha_1, \ldots, \alpha_m$ are finitely many vectors in V, each different from the zero vector, prove that there is a linear functional f on V such that

$$f(\alpha_i) \neq 0, \qquad i = 1, \ldots, m.$$

15. According to Exercise 3, similar matrices have the same trace. Thus we can define the trace of a linear operator on a finite-dimensional space to be the trace of any matrix which represents the operator in an ordered basis. This is well-defined since all such representing matrices for one operator are similar.

Now let V be the space of all 2×2 matrices over the field F and let P be a fixed 2×2 matrix. Let T be the linear operator on V defined by $T(A) = PA$. Prove that trace $(T) = 2$ trace (P).

16. Show that the trace functional on $n \times n$ matrices is unique in the following sense. If W is the space of $n \times n$ matrices over the field F and if f is a linear functional on W such that $f(AB) = f(BA)$ for each A and B in W, then f is a scalar multiple of the trace function. If, in addition, $f(I) = n$, then f is the trace function.

17. Let W be the space of $n \times n$ matrices over the field F, and let W_0 be the subspace spanned by the matrices C of the form $C = AB - BA$. Prove that W_0 is exactly the subspace of matrices which have trace zero. (*Hint:* What is the dimension of the space of matrices of trace zero? Use the matrix 'units,' i.e., matrices with exactly one non-zero entry, to construct enough linearly independent matrices of the form $AB - BA$.)

3.6. The Double Dual

One question about dual bases which we did not answer in the last section was whether every basis for V^* is the dual of some basis for V. One way to answer that question is to consider V^{**}, the dual space of V^*.

If α is a vector in V, then α induces a linear functional L_α on V^* defined by

$$L_\alpha(f) = f(\alpha), \qquad f \text{ in } V^*.$$

The fact that L_α is linear is just a reformulation of the definition of linear operations in V^*:

$$\begin{aligned} L_\alpha(cf + g) &= (cf + g)(\alpha) \\ &= (cf)(\alpha) + g(\alpha) \\ &= cf(\alpha) + g(\alpha) \\ &= cL_\alpha(f) + L_\alpha(g). \end{aligned}$$

If V is finite-dimensional and $\alpha \neq 0$, then $L_\alpha \neq 0$; in other words, there exists a linear functional f such that $f(\alpha) \neq 0$. The proof is very simple and was given in Section 3.5: Choose an ordered basis $\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$ for V such that $\alpha_1 = \alpha$ and let f be the linear functional which assigns to each vector in V its first coordinate in the ordered basis $\mathfrak{B}$.

Theorem 17. *Let* V *be a finite-dimensional vector space over the field* F. *For each vector* α *in* V *define*

$$\mathrm{L}_\alpha(\mathrm{f}) = \mathrm{f}(\alpha), \qquad \mathrm{f} \text{ in } \mathrm{V}^*.$$

The mapping $\alpha \to \mathrm{L}_\alpha$ *is then an isomorphism of* V *onto* V^{**}.

Proof. We showed that for each α the function L_α is linear. Suppose α and β are in V and c is in F, and let $\gamma = c\alpha + \beta$. Then for each f in V^*

$$\begin{aligned} L_\gamma(f) &= f(\gamma) \\ &= f(c\alpha + \beta) \\ &= cf(\alpha) + f(\beta) \\ &= cL_\alpha(f) + L_\beta(f) \end{aligned}$$

and so

$$L_\gamma = cL_\alpha + L_\beta.$$

This shows that the mapping $\alpha \to L_\alpha$ is a linear transformation from V into V^{**}. This transformation is non-singular; for, according to the remarks above $L_\alpha = 0$ if and only if $\alpha = 0$. Now $\alpha \to L_\alpha$ is a non-singular linear transformation from V into V^{**}, and since

$$\dim V^{**} = \dim V^* = \dim V$$

Theorem 9 tells us that this transformation is invertible, and is therefore an isomorphism of V onto V^{**}. ∎

Corollary. *Let* V *be a finite-dimensional vector space over the field* F. *If* L *is a linear functional on the dual space* V* *of* V, *then there is a unique vector* α *in* V *such that*

$$\mathrm{L}(\mathrm{f}) = \mathrm{f}(\alpha)$$

for every f *in* V*.

Corollary. *Let* V *be a finite-dimensional vector space over the field* F. *Each basis for* V* *is the dual of some basis for* V.

Proof. Let $\mathcal{B}^* = \{f_1, \ldots, f_n\}$ be a basis for V^*. By Theorem 15, there is a basis $\{L_1, \ldots, L_n\}$ for V^{**} such that

$$L_i(f_j) = \delta_{ij}.$$

Using the corollary above, for each i there is a vector α_i in V such that

$$L_i(f) = f(\alpha_i)$$

for every f in V^*, i.e., such that $L_i = L_{\alpha_i}$. It follows immediately that $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V and that $\mathcal{B}^*$ is the dual of this basis. ∎

In view of Theorem 17, we usually identify α with L_α and say that V 'is' the dual space of V^* or that the spaces V, V^* are naturally in duality with one another. Each is the dual space of the other. In the last corollary we have an illustration of how that can be useful. Here is a further illustration.

If E is a subset of V^*, then the annihilator E^0 is (technically) a subset of V^{**}. If we choose to identify V and V^{**} as in Theorem 17, then E^0 is a subspace of V, namely, the set of all α in V such that $f(\alpha) = 0$ for all f in E. In a corollary of Theorem 16 we noted that each subspace W is determined by its annihilator W^0. How is it determined? The answer is that W is the subspace annihilated by all f in W^0, that is, the intersection of the null spaces of all f's in W^0. In our present notation for annihilators, the answer may be phrased very simply: $W = (W^0)^0$.

Theorem 18. *If* S *is any subset of a finite-dimensional vector space* V, *then* $(\mathrm{S}^0)^0$ *is the subspace spanned by* S.

Proof. Let W be the subspace spanned by S. Clearly $W^0 = S^0$. Therefore, what we are to prove is that $W = W^{00}$. We have given one proof. Here is another. By Theorem 16

$$\dim W + \dim W^0 = \dim V$$
$$\dim W^0 + \dim W^{00} = \dim V^*$$

and since $\dim V = \dim V^*$ we have

$$\dim W = \dim W^{00}.$$

Since W is a subspace of W^{00}, we see that $W = W^{00}$. ∎

The results of this section hold for arbitrary vector spaces; however, the proofs require the use of the so-called Axiom of Choice. We want to avoid becoming embroiled in a lengthy discussion of that axiom, so we shall not tackle annihilators for general vector spaces. But, there are two results about linear functionals on arbitrary vector spaces which are so fundamental that we should include them.

Let V be a vector space. We want to define hyperspaces in V. Unless V is finite-dimensional, we cannot do that with the dimension of the hyperspace. But, we can express the idea that a space N falls just one dimension short of filling out V, in the following way:

1. N is a proper subspace of V;
2. if W is a subspace of V which contains N, then either $W = N$ or $W = V$.

Conditions (1) and (2) together say that N is a proper subspace and there is no larger proper subspace, in short, N is a maximal proper subspace.

Definition. *If* V *is a vector space, a* **hyperspace** *in* V *is a maximal proper subspace of* V.

Theorem 19. *If* f *is a non-zero linear functional on the vector space* V, *then the null space of* f *is a hyperspace in* V. *Conversely, every hyperspace in* V *is the null space of a (not unique) non-zero linear functional on* V.

Proof. Let f be a non-zero linear functional on V and N_f its null space. Let α be a vector in V which is not in N_f, i.e., a vector such that $f(\alpha) \neq 0$. We shall show that every vector in V is in the subspace spanned by N_f and α. That subspace consists of all vectors

$$\gamma + c\alpha, \qquad \gamma \text{ in } N_f, \; c \text{ in } F.$$

Let β be in V. Define

$$c = \frac{f(\beta)}{f(\alpha)}$$

which makes sense because $f(\alpha) \neq 0$. Then the vector $\gamma = \beta - c\alpha$ is in N_f since

$$\begin{aligned} f(\gamma) &= f(\beta - c\alpha) \\ &= f(\beta) - cf(\alpha) \\ &= 0. \end{aligned}$$

So β is in the subspace spanned by N_f and α.

Now let N be a hyperspace in V. Fix some vector α which is not in N. Since N is a maximal proper subspace, the subspace spanned by N and α is the entire space V. Therefore each vector β in V has the form

$$\beta = \gamma + c\alpha, \qquad \gamma \text{ in } N, \ c \text{ in } F.$$

The vector γ and the scalar c are uniquely determined by β. If we have also

$$\beta = \gamma' + c'\alpha, \qquad \gamma' \text{ in } N, \ c' \text{ in } F.$$

then

$$(c' - c)\alpha = \gamma - \gamma'.$$

If $c' - c \neq 0$, then α would be in N; hence, $c' = c$ and $\gamma' = \gamma$. Another way to phrase our conclusion is this: If β is in V, there is a unique scalar c such that $\beta - c\alpha$ is in N. Call that scalar $g(\beta)$. It is easy to see that g is a linear functional on V and that N is the null space of g. ∎

Lemma. *If* f *and* g *are linear functionals on a vector space* V, *then* g *is a scalar multiple of* f *if and only if the null space of* g *contains the null space of* f, *that is, if and only if* $\mathrm{f}(\alpha) = 0$ *implies* $\mathrm{g}(\alpha) = 0$.

Proof. If $f = 0$ then $g = 0$ as well and g is trivially a scalar multiple of f. Suppose $f \neq 0$ so that the null space N_f is a hyperspace in V. Choose some vector α in V with $f(\alpha) \neq 0$ and let

$$c = \frac{g(\alpha)}{f(\alpha)}.$$

The linear functional $h = g - \mathrm{cf}$ is 0 on N_f, since both f and g are 0 there, and $h(\alpha) = g(\alpha) - cf(\alpha) = 0$. Thus h is 0 on the subspace spanned by N_f and α—and that subspace is V. We conclude that $h = 0$, i.e., that $g = cf$. ∎

Theorem 20. *Let* g, $\mathrm{f}_1, \ldots, \mathrm{f}_r$ *be linear functionals on a vector space* V *with respective null spaces* N, $\mathrm{N}_1, \ldots, \mathrm{N}_r$. *Then* g *is a linear combination of* $\mathrm{f}_1, \ldots, \mathrm{f}_r$ *if and only if* N *contains the intersection* $\mathrm{N}_1 \cap \cdots \cap \mathrm{N}_r$.

Proof. If $g = c_1 f_1 + \cdots + c_r f_r$ and $f_i(\alpha) = 0$ for each i, then clearly $g(\alpha) = 0$. Therefore, N contains $N_1 \cap \cdots \cap N_r$.

We shall prove the converse (the 'if' half of the theorem) by induction on the number r. The preceding lemma handles the case $r = 1$. Suppose we know the result for $r = k - 1$, and let $f_1, \ldots, f_k$ be linear functionals with null spaces $N_1, \ldots, N_k$ such that $N_1 \cap \cdots \cap N_k$ is contained in N, the

null space of g. Let $g', f_1', \ldots, f_{k-1}'$ be the restrictions of $g, f_1, \ldots, f_{k-1}$ to the subspace N_k. Then $g', f_1', \ldots, f_{k-1}'$ are linear functionals on the vector space N_k. Furthermore, if α is a vector in N_k and $f_i'(\alpha) = 0$, $i = 1, \ldots, k-1$, then α is in $N_1 \cap \cdots \cap N_k$ and so $g'(\alpha) = 0$. By the induction hypothesis (the case $r = k - 1$), there are scalars c_i such that

$$g' = c_1 f_1' + \cdots + c_{k-1} f_{k-1}'.$$

Now let

$$h = g - \sum_{i=1}^{k-1} c_i f_i. \tag{3-16}$$

Then h is a linear functional on V and (3-16) tells us that $h(\alpha) = 0$ for every α in N_k. By the preceding lemma, h is a scalar multiple of f_k. If $h = c_k f_k$, then

$$g = \sum_{i=1}^{k} c_i f_i. \quad \blacksquare$$

Exercises

1. Let n be a positive integer and F a field. Let W be the set of all vectors $(x_1, \ldots, x_n)$ in F^n such that $x_1 + \cdots + x_n = 0$.

(a) Prove that W^0 consists of all linear functionals f of the form

$$f(x_1, \ldots, x_n) = c \sum_{j=1}^{n} x_j.$$

(b) Show that the dual space W^* of W can be 'naturally' identified with the linear functionals

$$f(x_1, \ldots, x_n) = c_1 x_1 + \cdots + c_n x_n$$

on F^n which satisfy $c_1 + \cdots + c_n = 0$.

2. Use Theorem 20 to prove the following. If W is a subspace of a finite-dimensional vector space V and if $\{g_1, \ldots, g_r\}$ is any basis for W^0, then

$$W = \bigcap_{i=1}^{r} N_{g_i}.$$

3. Let S be a set, F a field, and $V(S; F)$ the space of all functions from S into F:

$$(f + g)(x) = f(x) + g(x)$$
$$(cf)(x) = cf(x).$$

Let W be any n-dimensional subspace of $V(S; F)$. Show that there exist points $x_1, \ldots, x_n$ in S and functions $f_1, \ldots, f_n$ in W such that $f_i(x_j) = \delta_{ij}$.

3.7. The Transpose of a Linear Transformation

Suppose that we have two vector spaces over the field F, V, and W, and a linear transformation T from V into W. Then T induces a linear

transformation from W^* into V^*, as follows. Suppose g is a linear functional on W, and let

$$f(\alpha) = g(T\alpha) \tag{3-17}$$

for each α in V. Then (3-17) defines a function f from V into F, namely, the composition of T, a function from V into W, with g, a function from W into F. Since both T and g are linear, Theorem 6 tells us that f is also linear, i.e., f is a linear functional on V. Thus T provides us with a rule T^t which associates with each linear functional g on W a linear functional $f = T^t g$ on V, defined by (3-17). Note also that T^t is actually a linear transformation from W^* into V^*; for, if g_1 and g_2 are in W^* and c is a scalar

$$\begin{aligned}[T^t(cg_1 + g_2)](\alpha) &= (cg_1 + g_2)(T\alpha) \\ &= cg_1(T\alpha) + g_2(T\alpha) \\ &= c(T^t g_1)(\alpha) + (T^t g_2)(\alpha)\end{aligned}$$

so that $T^t(cg_1 + g_2) = cT^t g_1 + T^t g_2$. Let us summarize.

Theorem 21. *Let* V *and* W *be vector spaces over the field* F. *For each linear transformation* T *from* V *into* W, *there is a unique linear transformation* T^t *from* W* *into* V* *such that*

$$(\mathrm{T}^t\mathrm{g})(\alpha) = \mathrm{g}(\mathrm{T}\alpha)$$

for every g *in* W* *and* α *in* V.

We shall call T^t the **transpose** of T. This transformation T^t is often called the adjoint of T; however, we shall not use this terminology.

Theorem 22. *Let* V *and* W *be vector spaces over the field* F, *and let* T *be a linear transformation from* V *into* W. *The null space of* T^t *is the annihilator of the range of* T. *If* V *and* W *are finite-dimensional, then*

(i) *rank* (T^t) = *rank* (T)
(ii) *the range of* T^t *is the annihilator of the null space of* T.

Proof. If g is in W^*, then by definition

$$(T^t g)(\alpha) = g(T\alpha)$$

for each α in V. The statement that g is in the null space of T^t means that $g(T\alpha) = 0$ for every α in V. Thus the null space of T^t is precisely the annihilator of the range of T.

Suppose that V and W are finite-dimensional, say dim $V = n$ and dim $W = m$. For (i): Let r be the rank of T, i.e., the dimension of the range of T. By Theorem 16, the annihilator of the range of T then has dimension $(m - r)$. By the first statement of this theorem, the nullity of T^t must be $(m - r)$. But then since T^t is a linear transformation on an m-dimensional space, the rank of T^t is $m - (m - r) = r$, and so T and T^t have the same rank. For (ii): Let N be the null space of T. Every functional in the range

of T^t is in the annihilator of N; for, suppose $f = T^t g$ for some g in W^*; then, if α is in N

$$f(\alpha) = (T^t g)(\alpha) = g(T\alpha) = g(0) = 0.$$

Now the range of T^t is a subspace of the space N^0, and

$$\dim N^0 = n - \dim N = \text{rank } (T) = \text{rank } (T^t)$$

so that the range of T^t must be exactly N^0. ∎

Theorem 23. *Let* V *and* W *be finite-dimensional vector spaces over the field* F. *Let* $\mathcal{B}$ *be an ordered basis for* V *with dual basis* $\mathcal{B}^*$, *and let* $\mathcal{B}'$ *be an ordered basis for* W *with dual basis* $\mathcal{B}'^*$. *Let* T *be a linear transformation from* V *into* W; *let* A *be the matrix of* T *relative to* $\mathcal{B}$, $\mathcal{B}'$ *and let* B *be the matrix of* T^t *relative to* $\mathcal{B}'^*$, $\mathcal{B}^*$. *Then* $B_{ij} = A_{ji}$.

Proof. Let

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}, \qquad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\},$$
$$\mathcal{B}^* = \{f_1, \ldots, f_n\}, \qquad \mathcal{B}'^* = \{g_1, \ldots, g_m\}.$$

By definition,

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i, \qquad j = 1, \ldots, n$$

$$T^t g_j = \sum_{i=1}^{n} B_{ij} f_i, \qquad j = 1, \ldots, m.$$

On the other hand,

$$\begin{aligned}(T^t g_j)(\alpha_i) &= g_j(T\alpha_i) \\ &= g_j\left(\sum_{k=1}^{m} A_{ki}\beta_k\right) \\ &= \sum_{k=1}^{m} A_{ki} g_j(\beta_k) \\ &= \sum_{k=1}^{m} A_{ki}\delta_{jk} \\ &= A_{ji}.\end{aligned}$$

For any linear functional f on V

$$f = \sum_{i=1}^{m} f(\alpha_i) f_i.$$

If we apply this formula to the functional $f = T^t g_j$ and use the fact that $(T^t g_j)(\alpha_i) = A_{ji}$, we have

$$T^t g_j = \sum_{i=1}^{n} A_{ji} f_i$$

from which it immediately follows that $B_{ij} = A_{ji}$. ∎

Definition. *If* A *is an* m × n *matrix over the field* F, *the* **transpose** *of* A *is the* n × m *matrix* A^t *defined by* $A^t_{ij} = A_{ji}$.

Theorem 23 thus states that if T is a linear transformation from V into W, the matrix of which in some pair of bases is A, then the transpose transformation T^t is represented in the dual pair of bases by the transpose matrix A^t.

Theorem 24. *Let* A *be any* m × n *matrix over the field* F. *Then the row rank of* A *is equal to the column rank of* A.

Proof. Let $\mathfrak{B}$ be the standard ordered basis for F^n and $\mathfrak{B}'$ the standard ordered basis for F^m. Let T be the linear transformation from F^n into F^m such that the matrix of T relative to the pair $\mathfrak{B}$, $\mathfrak{B}'$ is A, i.e.,

$$T(x_1, \ldots, x_n) = (y_1, \ldots, y_m)$$

where

$$y_i = \sum_{j=1}^{n} A_{ij}x_j.$$

The column rank of A is the rank of the transformation T, because the range of T consists of all m-tuples which are linear combinations of the column vectors of A.

Relative to the dual bases $\mathfrak{B}'^*$ and $\mathfrak{B}^*$, the transpose mapping T^t is represented by the matrix A^t. Since the columns of A^t are the rows of A, we see by the same reasoning that the row rank of A (the column rank of A^t) is equal to the rank of T^t. By Theorem 22, T and T^t have the same rank, and hence the row rank of A is equal to the column rank of A. ∎

Now we see that if A is an $m \times n$ matrix over F and T is the linear transformation from F^n into F^m defined above, then

$$\text{rank } (T) = \text{row rank } (A) = \text{column rank } (A)$$

and we shall call this number simply the **rank** of A.

EXAMPLE 25. This example will be of a general nature—more discussion than example. Let V be an n-dimensional vector space over the field F, and let T be a linear operator on V. Suppose $\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$ is an ordered basis for V. The matrix of T in the ordered basis $\mathfrak{B}$ is defined to be the $n \times n$ matrix A such that

$$T\alpha_j = \sum_{j=1}^{n} A_{ij}\alpha_i$$

in other words, A_{ij} is the ith coordinate of the vector $T\alpha_j$ in the ordered basis $\mathfrak{B}$. If $\{f_1, \ldots, f_n\}$ is the dual basis of $\mathfrak{B}$, this can be stated simply

$$A_{ij} = f_i(T\alpha_j).$$

Let us see what happens when we change basis. Suppose

$$\mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$$

is another ordered basis for V, with dual basis $\{f_1', \ldots, f_n'\}$. If B is the matrix of T in the ordered basis $\mathcal{B}'$, then

$$B_{ij} = f_i'(T\alpha_j').$$

Let U be the invertible linear operator such that $U\alpha_j = \alpha_j'$. Then the transpose of U is given by $U^t f_i' = f_i$. It is easy to verify that since U is invertible, so is U^t and $(U^t)^{-1} = (U^{-1})^t$. Thus $f_i' = (U^{-1})^t f_i$, $i = 1, \ldots, n$. Therefore,

$$\begin{aligned} B_{ij} &= [(U^{-1})^t f_i](T\alpha_j') \\ &= f_i(U^{-1}T\alpha_j') \\ &= f_i(U^{-1}TU\alpha_j). \end{aligned}$$

Now what does this say? Well, $f_i(U^{-1}TU\alpha_j)$ is the i, j entry of the matrix of $U^{-1}TU$ in the ordered basis $\mathcal{B}$. Our computation above shows that this scalar is also the i, j entry of the matrix of T in the ordered basis $\mathcal{B}'$. In other words

$$\begin{aligned} [T]_{\mathcal{B}'} &= [U^{-1}TU]_{\mathcal{B}} \\ &= [U^{-1}]_{\mathcal{B}}[T]_{\mathcal{B}}[U]_{\mathcal{B}} \\ &= [U]_{\mathcal{B}}^{-1}[T]_{\mathcal{B}}[U]_{\mathcal{B}} \end{aligned}$$

and this is precisely the change-of-basis formula which we derived earlier.

Exercises

1. Let F be a field and let f be the linear functional on F^2 defined by $f(x_1, x_2) = ax_1 + bx_2$. For each of the following linear operators T, let $g = T^t f$, and find $g(x_1, x_2)$.

(a) $T(x_1, x_2) = (x_1, 0)$;
(b) $T(x_1, x_2) = (-x_2, x_1)$;
(c) $T(x_1, x_2) = (x_1 - x_2, x_1 + x_2)$.

2. Let V be the vector space of all polynomial functions over the field of real numbers. Let a and b be fixed real numbers and let f be the linear functional on V defined by

$$f(p) = \int_a^b p(x)\, dx.$$

If D is the differentiation operator on V, what is $D^t f$?

3. Let V be the space of all $n \times n$ matrices over a field F and let B be a fixed $n \times n$ matrix. If T is the linear operator on V defined by $T(A) = AB - BA$, and if f is the trace function, what is $T^t f$?

4. Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Let c be a scalar and suppose there is a non-zero vector α in V such that $T\alpha = c\alpha$. Prove that there is a non-zero linear functional f on V such that $T^t f = cf$.

5. Let A be an $m \times n$ matrix with *real* entries. Prove that $A = 0$ if and only if trace $(A^tA) = 0$.

6. Let n be a positive integer and let V be the space of all polynomial functions over the field of real numbers which have degree at most n, i.e., functions of the form

$$f(x) = c_0 + c_1x + \cdots + c_nx^n.$$

Let D be the differentiation operator on V. Find a basis for the null space of the transpose operator D^t.

7. Let V be a finite-dimensional vector space over the field F. Show that $T \to T^t$ is an isomorphism of $L(V, V)$ onto $L(V^*, V^*)$.

8. Let V be the vector space of $n \times n$ matrices over the field F.

(a) If B is a fixed $n \times n$ matrix, define a function f_B on V by $f_B(A) =$ trace (B^tA). Show that f_B is a linear functional on V.

(b) Show that every linear functional on V is of the above form, i.e., is f_B for some B.

(c) Show that $B \to f_B$ is an isomorphism of V onto V^*.